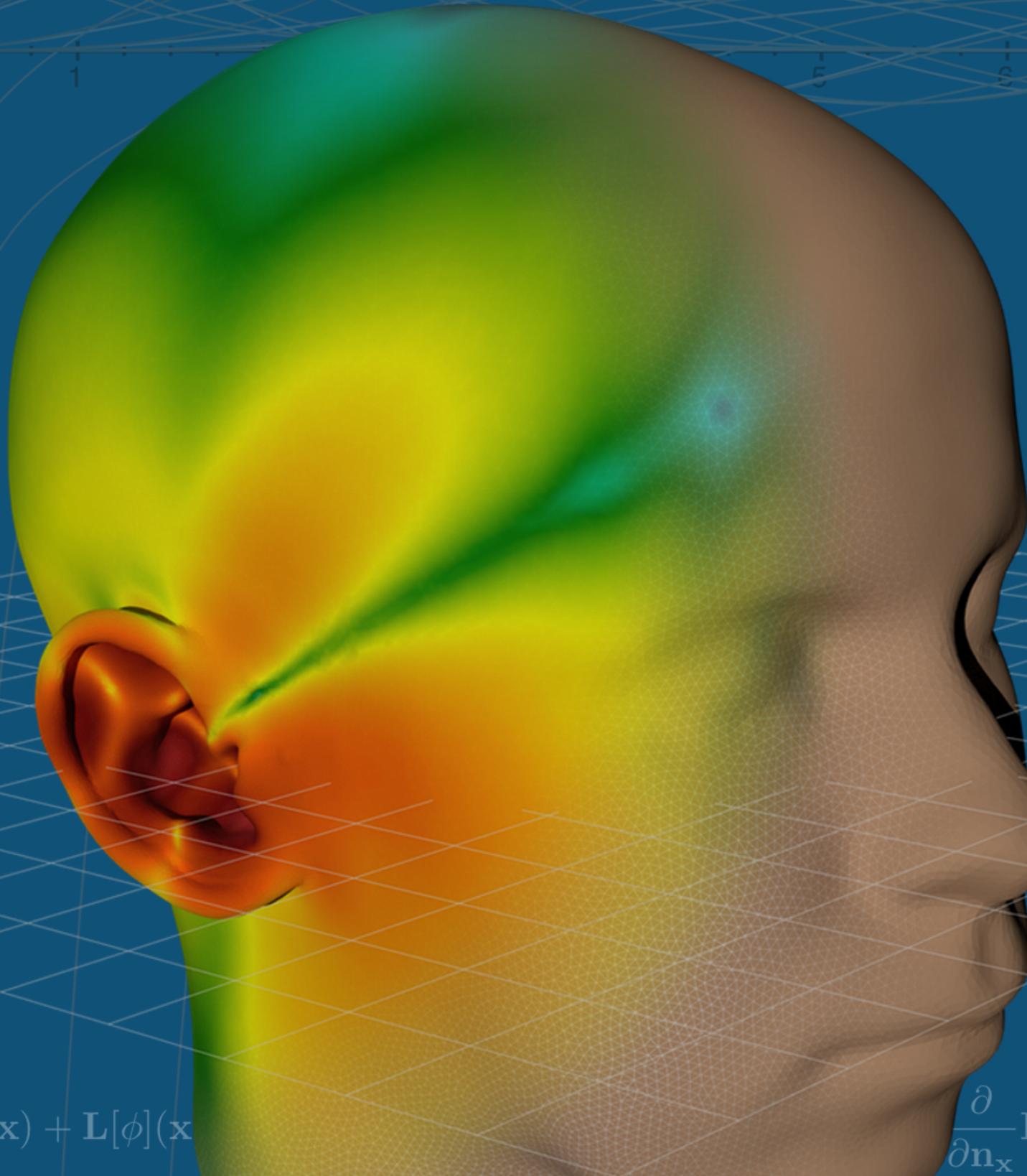


Numerical Calculation of Individual Head-Related Transfer Functions of Human Listeners

Harald Ziegelwanger



$$-\frac{1}{2}\phi(\mathbf{x}) + \mathbf{L}[\phi](\mathbf{x})$$

$$\frac{\partial}{\partial \mathbf{n}_x}$$



Numerical Calculation of Individual Head-Related Transfer Functions of Human Listeners

PhD Thesis

of

Harald Ziegelwanger

Date: 20. Mai 2016

Matriculation number: 0673086

PhD program: Sound and Music Computing (V 094 750)

Advisor: Prof. Dr. Robert Höldrich

Co-advisor: Prof. Dr. Gerhard Eckel

External reviewer: Prof. Dr.-Ing. Holger Waubke

Abstract

Head-related transfer functions (HRTFs) describe the directional filtering of the incoming sound at the ear canals. Spectral and temporal features of HRTFs are determined by the individual geometric details of a listener's head, torso and pinnae. Thus, listener-specific HRTFs are absolutely essential for an accurate sound localization in binaural 3-D audio reproduction systems. In contrast to the common acoustical measurements, where small microphones are placed at the entrances of the ear canals, HRTFs can be calculated numerically based on a listener's discretized geometry, captured by geometry acquisition techniques like laser-scanning or photogrammetric reconstruction. While acoustically measured HRTFs usually provide a sound-localization performance similar to that obtained in free-field listening, the performance obtained with numerically calculated HRTFs, however, depends on the quality of the geometric and acoustic model of the listener used for the numerical calculation. Apart from the problem of unclear requirements on the geometry, the computational effort of the numerical HRTF calculation is large and the calculation process lasts tens of hours.

The aim of this Phd project is to address open issues in the calculation of HRTFs under psychoacoustic quality criteria and to find methods to decrease the computational effort in the numerical HRTF calculation process. The modeling of the virtual microphone, the discretization of the boundary surface (head and ear) geometry and existing geometry-processing algorithms to reduce the computational effort in the BEM were investigated. Results were evaluated with a sagittal-plane sound-localization model, a time-of-arrival model and in sound-localization experiments. To stimulate further research in the field of binaural audio, the BEM code and the time-of-arrival model were published open-source as Mesh2HRTF and in the auditory modeling toolbox.

Kurzfassung

Außenohrübertragungsfunktionen (engl. head-related transfer functions, HRTFs) beschreiben die richtungsabhängige Filterung des einfallenden Schalls in den Ohrkanälen. Die spektralen und temporalen Merkmale von HRTFs sind bestimmt durch die individuellen geometrischen Details des Kopfes, des Torsos und vor allem der Außenohren. Deshalb sind hörspezifische HRTFs unumgänglich für eine akkurate Schallquellenlokalisierung in binauralen 3D-Audio-Systemen. Anders als in der üblichen akustischen Messung, in welcher kleine Mikrofone in den Ohrkanälen platziert werden, können HRTFs auch für ein diskretisiertes geometrisches Modell des Hörers, welches durch dreidimensionale Vermessungsverfahren wie Laser-Scanning oder photogrammetrische Rekonstruktion bestimmt werden kann, numerisch berechnet werden. Während akustisch gemessene HRTFs zu einer Lokalisationsleistung ähnlich wie beim Hören im Freifeld führen können, hängt die Lokalisationsleistung bei numerisch berechneten HRTFs stark von der Qualität des zugrundeliegenden geometrischen und akustischen Modell des Hörers ab. Abgesehen von den unklaren Anforderungen an die Geometrie ist der Rechenaufwand dieser Methode enorm und die Berechnung dauert mehrere Stunden.

Das Ziel dieser Doktorarbeit ist die Klärung offener Fragen in der numerischen Berechnung von HRTFs unter psychoakustischen Qualitätskriterien und das Finden von Methoden zur Reduzierung der Rechenaufwandes im Prozess der numerischen HRTF-Berechnung. Der Effekt des Mikrofonmodells, die Auswirkung der Diskretisierung der Geometrie und die Anwendbarkeit existierender Algorithmen zur Optimierung von Gitternetzen zur Reduzierung des Rechenaufwandes in der BEM wurden untersucht. Ergebnisse wurden mit einem Lokalisationsmodell für Schallquellen in Sagittalebene, mit einem Modell der Signallaufzeiten und in Lokalisationsexperimenten evaluiert. Als Motivation für weiterführende Untersuchungen in Bereich der Binauraltechnik wurde der BEM-Code als Mesh2HRTF und das Modell der Signallaufzeiten in der Auditory Modeling Toolbox quelloffen veröffentlicht.



(Name in Blockbuchstaben)

(Matrikelnummer)

Erklärung

Hiermit bestätige ich, dass mir der *Leitfaden für schriftliche Arbeiten an der KUG* bekannt ist und ich diese Richtlinien eingehalten habe.

Graz, den

.....
Unterschrift der Verfasserin/des Verfassers

Acknowledgements

I worked on this PhD thesis at the Acoustics Research Institute (ARI) of the Austrian Academy of Sciences in the context of the FWF-funded project LocaPhoto. So first and foremost, I offer my sincerest gratitude to Wolfgang Kreuzer and Piotr Majdak, who have supported me throughout my PhD project. My sincere thanks also go to all colleagues from the ARI – particularly to members of the psychoacoustics group and the computational acoustics group for lots of fruitful discussions and for the friendly working atmosphere. I want to thank Michael Mihocic for his assistance with measuring head-related transfer functions and with conducting sound-localization experiments.

I am also very grateful to Robert Höldrich and Gerhard Eckel from the Institute of Electronic Music and Acoustics at the University of Music and Performing Arts Graz (KUG) for advising my PhD project from the university side. Moreover, I want to thank the doctoral school at KUG for administrative support and granting my work. My sincere thank also goes to the International Commission on Acoustics, the European Acoustics Association and the German Acoustics Association for various grants enabling me to present my work at international conferences and exchanging my findings and ideas with experts from the field.

Most importantly, none of this would have been possible without the patience and support of my family. I would like to express my heartfelt gratitude to my mum. Words can not express how grateful I am for all of the sacrifices that you've made on my behalf. My deepest gratitude goes to my beloved wife Julia who experienced all of the ups and downs of my research. You have been a true and great supporter and have unconditionally loved me during my good and bad times. I can't thank you enough for encouraging me throughout this experience. Finally, I am thankful to my beloved cat Miriel for cheering me up during the last three years.

Harald Ziegelwanger, Krems an der Donau

Contents

Abstract	III
Acknowledgements	VII
Contents	IX
1 Introduction	1
2 Mesh2HRTF: An open-source software package for the numerical calculation of HRTFs	7
3 Modeling the direction-continuous time-of-arrival in HRTFs	17
4 Calculation of listener-specific HRTFs: Effect of mesh quality	35
5 Numerical calculation of listener-specific HRTFs and sound localization: Microphone model and mesh discretization	45
6 Non-uniform sampling of geometry for the numeric simulation of HRTFs	61
7 A-priori mesh grading for the numerical calculation of the HRTFs	71
8 Efficient numerical calculation of HRTFs	85
9 Concluding remarks	93
Bibliography	97

Chapter 1

Introduction

The auditory system allows human listeners to localize sound sources in three-dimensional (3D) space with just two ears (Blauert, 1997). In the localization process, the auditory system evaluates acoustic features of the incoming sound. Different features are relevant to determine the lateralization, elevation, and distance of a sound source, respectively. For the lateralization, interaural differences play a major role, viz., the interaural time difference (ITD) and the interaural level difference (ILD). While the ITD is the dominant feature at low frequencies and when lateralizing broadband sound, the ILD is important for frequencies above 1.5 kHz (Macpherson and Middlebrooks, 2002). The elevation perception of a sound source and the front-back discrimination is based on the evaluation of spectral features (Middlebrooks and Green, 1990; Middlebrooks, 1999a,b; Bronkhorst, 1995). For the distance perception, intensity and spectral cues are relevant (Coleman, 1963; Mershon and King, 1975).

The acoustic features can be described by the so-called head-related transfer functions (HRTFs, Mehrgardt and Mellert, 1977; Wightman and Kistler, 1989a,b). In particular, they describe the filtering of the incoming sound at a listener's ear-canals (Xie, 2013). The filtering is caused by reflection, diffraction, shadowing and scattering at a listener's anatomical structure, i.e., the head, torso, and particularly the pinnae. The shadowing, diffraction and scattering around the head result in interaural differences. Reflections at the fine geometry of a listener's pinnae result in spectral features, e.g., peaks and notches in the HRTF amplitude spectra (Takemoto *et al.*, 2012). Torso reflections also generate spectral features at low frequencies (Algazi *et al.*, 2001), however, these features play a minor role in the sound-localization process (Carlile *et al.*, 1999). With a complete set of HRTFs, arbitrary combinations of virtual sound sources can be presented via headphones, viz., virtual auditory display (Xie, 2013), a procedure known as binaural reproduction technique (Møller, 1992). However, such systems often use non-individual

or generic HRTFs and, thus, are not able to provide accurate listener-specific spatial information. The result can be a diffuse sound source position, an incorrect localization (Wenzel *et al.*, 1993), and front-back confusion of sound sources. Non-individual HRTFs can even lead to perceiving a sound source within the head along the interaural axis rather than externalizing it at the actual position (Hartmann and Wittenberg, 1996). Hence, for an accurate binaural 3D sound reproduction, listener-specific HRTFs are absolutely essential.

Listener-specific HRTFs are usually acquired by acoustic measurement (Wightman and Kistler, 1989a,b; Møller *et al.*, 1995; Majdak *et al.*, 2007; Dietrich *et al.*, 2013). However, the acoustic measurement is technically complex, requires dedicated equipment and measurement facilities, includes the placement of small microphones in a listener’s ear canals, and lasts a substantial amount of time while the listener is not allowed to move. As an alternative to the acoustical measurement procedure, the acquisition of a listener’s morphology and its processing as a 3D polygon mesh, i.e., a set of nodes and triangular or rectangular elements, provides the basis to calculate HRTFs using numerical methods, i.e., by simulating the sound field of an incident wave scattered by a listener’s head and pinnae. Different geometry acquisition techniques and numerical methods were tested for the numerical calculation of HRTFs. For the geometry acquisition, laser scanning (Heritage and Large, 2009; Lu-Xingchang and Liu-Xianlin, 2006), X-ray computed tomography (CT-scan, Herman, 2009), magnetic resonance imaging (MRI, Brown *et al.*, 2014), and photogrammetric reconstruction (Pears *et al.*, 2012) were used in Katz (2001a), Gumerov *et al.* (2010), Jin *et al.* (2014), and Rébillat *et al.* (2014), respectively. For the numerical calculation of HRTFs, the finite-difference time domain method (FDTD, Yee, 1966) was used in Takemoto *et al.* (2012), the ultra-weak variational formulation (Cessenat and Despres, 1998) was used in Huttunen *et al.* (2007), the finite element method (FEM, Zienkiewicz *et al.*, 2013) was used in Ma *et al.* (2015) and the boundary element method (BEM, Brebbia and Dominguez, 1977; Ciskowski and Brebbia, 1991) was used for the first time in Katz (2001a,b) to calculate HRTFs numerically. Because of the computational effort of the BEM, the first attempt to calculate HRTFs, however, was limited to 22 000 elements in the mesh and, thus, to frequencies below 5.4 kHz (Katz, 2001a,b). The numerical HRTF calculation became feasible for the full audible frequency range (Kreuzer *et al.*, 2009; Gumerov *et al.*, 2010) by coupling the BEM with the fast multipole method (FMM, Greengard and Rokhlin, 1987).

The quality of the numerically calculated HRTFs depends on the quality of the mesh and on the accuracy of the numerical algorithms used for the calculation. The quality of the mesh mainly depends on the geometry acquisition technique used to capture the geometry of a listener’s head and pinna. While HRTFs calculated for a CT-scan of a KEMAR

mannequin (Gumerov *et al.*, 2010) were very similar to acoustically measured HRTFs when comparing HRTF amplitude spectra visually, HRTFs calculated for a laser-scan of a real human subject (Kreuzer *et al.*, 2009) showed obvious visual differences in the amplitude spectra. The numerical accuracy of the BEM mainly depends on the number of elements in the mesh used for the calculation. In Marburg (2002) the relative numerical error of BEM calculations was below fifteen percent, when at least six elements per wavelength and a uniform element size were used. For the numerical calculation of HRTFs, Katz (2001a) suggested six elements per wavelength and a homogenous geometry discretization and Gumerov *et al.* (2010) recommended five elements per wavelength, equilateral triangles, and a uniform vertex distribution in the mesh. Nevertheless, requirements on the numerical calculation of listener-specific HRTFs which yield a good sound-localization performance have not been clarified. While acoustically measured HRTFs were validated in sound-localization experiments (Middlebrooks and Green, 1990; Middlebrooks, 1999a,b; Bronkhorst, 1995; Majdak *et al.*, 2010), the evaluation of numerically calculated HRTFs was based on the comparison of HRTF amplitude spectra only (Katz, 2001a; Kahana and Nelson, 2006, 2007; Gumerov *et al.*, 2010; Kreuzer *et al.*, 2009).

This PhD project focused on the perception-based evaluation of numerically calculated HRTFs. To evaluate HRTFs for a large set of conditions, numerically calculated HRTFs were evaluated by perception-based models. These models were used to compare temporal features in an HRTF set and to predict the listener-specific sound-localization performance in sagittal planes yielded by an HRTF set. The most relevant conditions were then validated in a sound-localization experiment. Another focus was the reduction of the computational effort of the numerical HRTF calculation. The following chapters contain articles that describe the numerical calculation method, the models used to evaluate calculation results, the evaluation of requirements for the numerical calculation of listener-specific HRTFs and methods to reduce the computational effort in the process of the HRTF calculation. Chapter 2 starts with the description of the software package used for the numerical calculation of HRTFs, viz., *Mesh2HRTF* (chapter 2). The software package was developed based on an existing BEM code (Acoustics Research Institute, OEAW, Chen *et al.*, 2008) and published under the *GNU Lesser General Public License*. The major requirement for the software package was to reduce the number of manual processing steps in the calculation process to enable the numerical HRTF calculation of a large set of conditions in a reasonable time. In this PhD thesis, HRTFs for approximately 500 meshes were numerically calculated with *Mesh2HRTF*.

Since the perceptual evaluation of HRTFs in sound-localization experiments is time intensive, models can be used to quickly compare and evaluate HRTFs, e.g., the temporal and spectral features, for a large number of conditions. For the comparison of spectral

features in HRTFs, an existing sagittal plane sound-localization model, which bases on [Langendijk and Bronkhorst \(2002\)](#) and was further developed and evaluated in [Baumgartner et al. \(2013, 2014\)](#) and [Majdak et al. \(2014\)](#) was used in this thesis. For the comparison and evaluation of temporal features, a model of the time-of-arrival (TOA, see [chapter 3](#)) was developed, implemented and evaluated. The work was initiated in the diploma thesis of the author, where it was shown that a simplified geometrical model of the head can be used to estimate a listener's head radius and to model the listener-specific TOA and ITD. The development of that model was continued, and the effect of a pinna and torso on the temporal features was investigated. The investigation based on numerically calculated HRTFs of a simple rigid sphere, a sphere model which also included a generic pinna, and the same sphere and pinna model which additionally included a torso.

While the numerical HRTF calculation for meshes in [chapter 3](#) were based on CAD models, meshes of real human subjects had to be captured by the use of a geometry acquisition technique for the investigation of the numerical calculation of listener-specific HRTFs. The geometry of three human subjects was captured and processed as polygon mesh by a project partner (Institute for Virtual Reality and Visualization, VRVis). The geometry acquisition technique based on CT-scans of subjects' pinnae silicone imprints is described in [chapter 4](#) where also the effect of the geometry discretization was preliminarily evaluated for HRTFs of a single subject. Results showed, that the modeled sound-localization performance degraded when the size of the elements in the mesh was increased.

In the subsequent study ([chapter 5](#)), the different errors arising in the process of numerically calculating HRTFs were investigated, viz., geometrical, numerical, and acoustical errors. The geometric error arises because of under-sampling of the geometry, the numerical error arises because of under-sampling of the sound field on the geometry, and acoustical errors arise when simplifying the physical reality in models, e.g., the microphone model. The effects of the different error types were separated and relevant parameters required for perceptually valid numerically calculated HRTFs in terms of a good sound-localization performance were identified. The effect of the virtual microphone and the effect of the mesh resolution were systematically evaluated. Both effects were evaluated using model predictions. For the most relevant conditions, numerically calculated HRTFs were finally verified in a sound-localization experiment. Results showed that approximately 80 000 elements were required to calculate HRTFs yielding the same localization performance as acoustically measured HRTFs. As a drawback, the computation time for such a number of elements was in the order of hours. Reducing the computation time by a simple mesh coarsening considering all elements, i.e., a *uniform* re-meshing, also reduced the sound-localization performance.

In other fields of computational physics, more sophisticated geometry discretization methods, viz., goal-oriented mesh adaptivity (Walsh and Demkowicz, 2003; Chen *et al.*, 2002), non-uniform meshes (Goldstein, 1982), and mesh grading (Heinrich *et al.*, 1996), resulting in *non-uniform* meshes, had been investigated. Thus, the effect of non-uniform meshes on the quality of numerically calculated HRTFs and on the the computational effort in the calculations was investigated. A rough approximation of the potential of reducing the computational effort by non-uniform sampling of the geometry was assessed in [chapter 6](#). Here, HRTFs were numerically calculated for a non-uniform mesh, which was generated by an existing adaptive re-meshing algorithm. In this algorithm, the length of element edges was varied as a function of the smoothness of the geometry.

The subsequent study ([chapter 7](#)) continued on non-uniform re-meshing and seized on the concept of a-priori mesh grading (Langer *et al.*, 2015). An a-priori mesh grading algorithm was developed and implemented based on an existing uniform re-meshing algorithm (Botsch and Kobbelt, 2004). Different grading functions were tested and the numerical and perceptual error in HRTFs calculated for a-priori graded meshes was evaluated for different geometrical objects. The a-priori mesh grading reduced the number of elements in the mesh to approximately 13 000 elements and the computation time approximately by a factor of seven while keeping the numerical and perceptual error at the same level as for uniform meshes with approximately 80 000 elements. It turned out that the FMM was inefficient when calculating HRTFs in the far-field. Thus, in [chapter 8](#) HRTFs were additionally calculated in the near-field and extrapolated to the far-field by range extrapolation (Pollow *et al.*, 2012), which further reduced computation time in the HRTF calculation process.

In [chapter 9](#), the general results and findings of this PhD project are summarized and an outlook on further investigations is given.

Chapter 2

Mesh2HRTF: An open-source software package for the numerical calculation of HRTFs

This work was published as:

Ziegelwanger, H., Majdak, P., and Kreuzer, W. (2015). "Mesh2HRTF: An open-source software package for the numerical calculation of head-related transfer functions," in *Proceedings of the 22nd International Congress of Sound and Vibration*, Florence, IT, 1-8. [doi:10.13140/RG.2.1.1707.1128](https://doi.org/10.13140/RG.2.1.1707.1128)

The original BEM-code was developed at the Acoustics Research Institute by Zhensheng Chen. Based on the initial idea of the second author to use the BEM for the calculation of HRTFs, I adapted the BEM-code for the LocaPhoto project and for the publication as open-source software, added the pre- and post-processing modules, and evaluated the code in numerical experiments. I also generated the figures and wrote the initial draft of the manuscript. The co-authors provided feedback for the design of the software package and helped writing the manuscript.

Co-author 1:

Co-author 2:



MESH2HRTF: AN OPEN-SOURCE SOFTWARE PACKAGE FOR THE NUMERICAL CALCULATION OF HEAD-RELATED TRANSFER FUNCTIONS

Harald Ziegelwanger, Wolfgang Kreuzer, Piotr Majdak

*Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria
email: harald.ziegelwanger@oeaw.ac.at*

Head-related transfer functions (HRTFs) describe the listener-specific directional filtering of the incoming sound at the ear-canal. HRTFs are usually obtained in an acoustic measurement procedure. However, they can also be numerically calculated based on a geometric representation of listeners' head and pinnae. Numerical methods like the fast multipole method and mesh optimization techniques like mesh-grading enable feasible the efficient numerical calculation of HRTFs with the boundary element method (BEM) for the whole human audible frequency range. In this paper, MESH2HRTF is presented. MESH2HRTF is an open-source project aiming at providing an easy-to-use software package for the numerical calculation of HRTFs. It targets researchers in the field of binaural audio. The core of MESH2HRTF is an implementation of the 3-dimensional Burton-Miller collocation BEM coupled with the multi-level fast multipole method. MESH2HRTF also includes tools for the preprocessing of geometry data, for the generation of evaluation grids, and for the post-processing of calculation results. Results are stored in the spatially oriented format for acoustics (SOFA). MESH2HRTF has already been used for the numerical calculation of perceptually valid listener-specific HRTFs. In this paper, MESH2HRTF was evaluated numerically by means of comparing the numerical and the analytical solution for a sphere.

1. Introduction

Head-related transfer functions (HRTFs) describe the listener-specific directional filtering of the incoming sound at the entrance of the ear-canal [1, 2]. Since HRTFs are caused by a scattering process, it seems natural to calculate HRTFs numerically by simulating the sound-field scattered by the human head and ear instead of acquiring HRTFs in an acoustical measurement procedure. Different simulation methods for the numerical calculation of HRTFs were investigated [3, 4, 5] and particularly much research effort has been put into the boundary element method (BEM) [6]. HRTFs were numerically calculated with the BEM and evaluated in several studies [3, 7, 8, 9, 10, 11, 12, 13, 14]. In most of these studies, researchers either had to use commercial software or to implement the BEM by themselves. Both cases make the access to the numerical calculation of HRTFs and reproducible research difficult.

Thus, in this paper, MESH2HRTF is presented. MESH2HRTF is an open-source project¹ aiming at providing an easy-to-use software package for the numerical calculation of HRTFs. It targets researchers in the field of binaural audio. In a nutshell, MESH2HRTF simply reads geometrical data, calculates the corresponding sound field and outputs HRTFs. To support multiple computer platforms,

¹ Available from <http://mesh2hrtf.sourceforge.net>

the concept of MESH2HRTF is to focus on a command-line tool, which forms the numerical core (see Sec. 3.2), i.e., an implementation of the 3-dimensional Burton-Miller collocation BEM coupled with the multi-level fast multipole method (ML-FMM), and to provide add-ons for existing cross-platform applications for the preprocessing of geometry and for the visualization of results.

MESH2HRTF was used for the numerical calculation of HRTFs in several studies [10, 15, 13, 14, 16, 17]. Results were evaluated numerically and perceptually. In the following, an overview on the theoretical background and on the functionality of MESH2HRTF is given.

2. Theoretical Background

2.1 Boundary integral equation

In general, the Helmholtz equation $\nabla^2 \phi(\mathbf{x}) + k^2 \phi(\mathbf{x}) = 0$ describes acoustic waves in a domain Ω_e , i.e., Ω_e is the exterior domain outside an object Ω . $\phi(\mathbf{x}) = \frac{p(\mathbf{x})}{i\omega\rho}$ denotes the velocity potential at the point \mathbf{x} , and $k = \frac{\omega}{c}$ is the wavenumber dependent on the speed of sound c and the circular frequency ω . $p(\mathbf{x})$ represents the sound pressure at a point \mathbf{x} , ρ the density of the medium and i the imaginary unit. \mathbf{n} is the vector normal to the object's surface Γ at a point \mathbf{x} pointing to Ω_e and $v(\mathbf{x}) = \frac{\partial \phi(\mathbf{x})}{\partial \mathbf{n}}$ is the particle velocity in direction of \mathbf{n}_x . For the BEM, the Helmholtz equation is transformed into a boundary integral equation (BIE):

$$(1) \quad -\frac{1}{2}\phi(\mathbf{x}) + \int_{\Gamma} H(\mathbf{x}, \mathbf{y})\phi(\mathbf{y})d\mathbf{y} - \int_{\Gamma} G(\mathbf{x}, \mathbf{y})v(\mathbf{y})d\mathbf{y} = -\phi_{inc}(\mathbf{x}), \quad \mathbf{x} \in \Gamma.$$

By taking the derivative of Eq. (1) with respect to \mathbf{n}_x and by multiplying this derivative with $\frac{i}{k}$

$$(2) \quad -\frac{i}{2k}v(\mathbf{x}) + \frac{i}{k} \int_{\Gamma} E(\mathbf{x}, \mathbf{y})\phi(\mathbf{y})d\mathbf{y} - \frac{i}{k} \int_{\Gamma} H^T(\mathbf{x}, \mathbf{y})v(\mathbf{y})d\mathbf{y} = -\frac{i}{k}v_{inc}(\mathbf{x}), \quad \mathbf{x} \in \Gamma,$$

the Burton-Miller method [18] is applied by taking the sum of Eq. (1) and Eq. (2). $G(\mathbf{x}, \mathbf{y}) = \frac{e^{ik\|\mathbf{y}-\mathbf{x}\|}}{4\pi\|\mathbf{y}-\mathbf{x}\|}$, $H(\mathbf{x}, \mathbf{y}) = \frac{\partial}{\partial \mathbf{n}_y}G(\mathbf{x}, \mathbf{y})$, $H^T(\mathbf{x}, \mathbf{y}) = \frac{\partial}{\partial \mathbf{n}_x}G(\mathbf{x}, \mathbf{y})$, and $E(\mathbf{x}, \mathbf{y}) = \frac{\partial}{\partial \mathbf{n}_x} \frac{\partial}{\partial \mathbf{n}_y}G(\mathbf{x}, \mathbf{y})$ are the Green's function of the Helmholtz equation and its derivatives, respectively.

In the most general way, boundary conditions are described by

$$(3) \quad \alpha \frac{\partial \phi(\mathbf{x})}{\partial \mathbf{n}} + \beta k^2 \phi(\mathbf{x}) = f, \quad \mathbf{x} \in \Gamma,$$

which corresponds to *Robin* boundary conditions. *Dirichlet* boundary conditions can be described by $\alpha = 0$, e.g., $\alpha = 0$ and $f = 0$ represents a sound-soft surface. *Neumann* boundary conditions can be described by $\beta = 0$, e.g., $\beta = 0$ and $f = 0$ represents a sound-hard surface. For the reciprocal calculation of HRTFs, an active vibrating surface area, i.e., the area of the virtual microphone, can be defined by setting $\beta = 0$ and $f = v_{mic}$ at the microphone area Γ_{mic} .

2.2 Collocation

For numerical calculations, the boundary surface Γ has to be discretized, e.g., as a 3-dimensional polygon mesh \mathcal{M} . \mathcal{M} consists of vertices $\mathcal{V} = \{v_1, \dots, v_j, \dots, v_J\}$, edges $\mathcal{E} = \{e_1, \dots, e_k, \dots, e_K\}$ and faces $\mathcal{F} = \{f_1, \dots, f_n, \dots, f_N\}$. Γ_{f_n} is the surface of an element. The collocation method with piecewise constant basis functions discretizes the Burton-Miller formulation of the BIE at collocation nodes \mathbf{x}_m (\mathbf{x}_m is the midpoint of the face f_m). The resulting linear system of equations is defined as:

$$(4) \quad \mathbf{A}_1 \boldsymbol{\phi} + \mathbf{A}_2 \mathbf{v} = \boldsymbol{\phi}_{inc} + \frac{i}{k} \mathbf{v}_{inc}.$$

where $\boldsymbol{\phi}_m = \boldsymbol{\phi}(\mathbf{x}_m)$, $\mathbf{v}_m = \mathbf{v}(\mathbf{x}_m)$, $\boldsymbol{\phi}_m^{inc} = \boldsymbol{\phi}_{inc}(\mathbf{x}_m)$, $\mathbf{v}_m^{inc} = \mathbf{v}_{inc}(\mathbf{x}_m)$,

$$(5) \quad \begin{aligned} A_{m,n}^1 &= -\frac{1}{2} \delta_{m,n} + \int_{\Gamma_{fn}} H(\mathbf{x}_m, \mathbf{y}) d\mathbf{y} + \frac{i}{k} \int_{\Gamma_{fn}} E(\mathbf{x}_m, \mathbf{y}) d\mathbf{y}, \\ A_{m,n}^2 &= -\frac{i}{2k} \delta_{m,n} - \int_{\Gamma_{fn}} G(\mathbf{x}_m, \mathbf{y}) d\mathbf{y} - \frac{i}{k} \int_{\Gamma_{fn}} H^T(\mathbf{x}_m, \mathbf{y}) d\mathbf{y}. \end{aligned}$$

2.3 Multipole expansion

If Γ is clustered (see Sec. 3.2.1) and the distance of the clusters is sufficiently large then the integral kernels from Eq. (1) and Eq. (2) can be approximated by the multipole expansion, e.g., G can be expanded to:

$$(6) \quad \begin{aligned} G(\mathbf{x}, \mathbf{y}) &= \frac{ik}{4\pi} \sum_{l=0}^{\infty} (2l+1) i^l h_l^{(1)}(k\|\mathbf{r}_0\|) \int_{\mathcal{S}} e^{iks(\mathbf{x}-\mathbf{z}_2-\mathbf{y}+\mathbf{z}_1)} P_l\left(\frac{\mathbf{s} \cdot \mathbf{r}_0}{\|\mathbf{r}_0\|}\right) ds \\ &= \lim_{L \rightarrow \infty} \frac{ik}{4\pi} \int_{\mathcal{S}} e^{iks(\mathbf{x}-\mathbf{z}_2)} M_L(\mathbf{s}, \mathbf{r}_0) e^{-iks(\mathbf{y}-\mathbf{z}_1)} ds, \end{aligned}$$

where

$$(7) \quad M_L(\mathbf{s}, \mathbf{r}_0) := \sum_{l=0}^L (2l+1) i^l h_l^{(1)}(k\|\mathbf{r}_0\|) P_l\left(\frac{\mathbf{s} \cdot \mathbf{r}_0}{\|\mathbf{r}_0\|}\right).$$

\mathbf{z}_1 and \mathbf{z}_2 are the midpoints of two clusters, $\mathbf{r}_0 := \mathbf{z}_2 - \mathbf{z}_1$, \mathcal{S} is the unit-sphere, $h_l^{(1)}(\cdot)$ are the spherical Hankel functions of the first kind and order l , and $P_l(\cdot)$ are the Legendre polynomials of order l . L is the truncation order. Similar expansions exist for $H(\mathbf{x}, \mathbf{y})$, $H^T(\mathbf{x}, \mathbf{y})$, and $E(\mathbf{x}, \mathbf{y})$. For more details refer to Ref. [19] and [20].

3. Software Package

3.1 Input Files

3.1.1 Boundary Mesh

At the moment, Ω is limited to single component, closed, orientable manifolds and \mathcal{M} is limited to triangular polygon meshes. BLENDER² is recommended to generate and evaluate polygon meshes. For the preprocessing of geometrical data OPENFLIPPER³ is recommended, i.e., a cross-platform open-source application designed for processing, modeling and rendering of geometric data [21]. OPENFLIPPER offers the functionality to apply different mesh manipulation techniques, e.g., isotropic re-meshing, adaptive re-meshing, and mesh-simplification. MESH2HRTF includes a dedicated re-meshing add-on for the preprocessing and optimization of geometrical data for the BEM [16, 17].

While MESH2HRTF provides its own proprietary file format, i.e., an VTK like ASCII format, MESH2HRTF also reads boundary meshes in common file formats, e.g., OBJ, STL, and PLY, by the use of OPENMESH⁴ [22], which is an open-source, generic and efficient polygon mesh data structure.

² Available from <http://www.blender.org> (date last viewed: March 13, 2015)

³ Available from <http://www.openflipper.org> (date last viewed: March 13, 2015)

⁴ Available from <http://www.openmesh.org> (date last viewed: March 13, 2015)

3.1.2 Evaluation Grid

MESH2HRTF provides add-ons for BLENDER to generate and to export evaluation grids optimized for Ω , e.g., hyper-interpolation grids [23] on the bounding sphere of Ω allow an efficient calculation of HRTFs in the far-field by near-field calculation and range extrapolation in the modal domain [17].

3.1.3 Settings

The settings file contains information about the temperature of the medium, the density of the medium, the speed of sound, parameters (position, source strength) for point sources, parameters for plane waves, frequencies to be evaluated, and boundary conditions.

3.2 Numerical Core

3.2.1 Stage 1: Building

In the simplest case, the numerical core builds the linear system of equations:

$$(8) \quad \mathbf{A}\hat{\mathbf{x}} = \mathbf{b}.$$

Dependent on the boundary condition (Eq. 3), $\hat{\mathbf{x}}$ consists of entries for ϕ or \mathbf{v} , \mathbf{A} is defined by \mathbf{A}_1 and/or \mathbf{A}_2 , and \mathbf{b} is defined by the incident field of the external sound sources, i.e., ϕ_{inc} and \mathbf{v}_{inc} , and velocity and pressure boundary conditions. For computing the entries of \mathbf{A} , the integral kernels for all combinations of collocation nodes \mathbf{x}_m and field elements f_n are integrated over Γ_{f_n} . If \mathbf{x}_m is not located on f_n regular quadrature is used to compute the integral and singular quadrature otherwise.

Regular quadrature: The integration is done by Gauss quadrature over the unit-triangle and the result is then transformed to the real element by using the Jacobian. When \mathbf{x}_m is located close to the field element f_n , f_n is non-uniformly subdivided in order to enhance numerical stability. As a criterion for the subdivision $\frac{\|\mathbf{x}_m - \mathbf{c}_n\|}{\sqrt{A_n}} < 1.3$ is used, where \mathbf{c}_n is the midpoint and A_n is the area of a (sub-)element.

Singular quadrature: Three types of singularity occur in Eq. (4), i.e. weak-singularity from $G(\mathbf{x}, \mathbf{y})$, singularity from $H(\mathbf{x}, \mathbf{y})$ and $H^T(\mathbf{x}, \mathbf{y})$, and hyper-singularity from $E(\mathbf{x}, \mathbf{y})$. Weak-singular and singular integrals are calculated by using the duffy transformation [24]. Hyper-singular integrals are converted into regular line and surface integrals using Stokes theorem [25].

\mathbf{A} is a densely populated and unsymmetrical $N \times N$ matrix. If the number of elements N is large, it is necessary to use the FMM in order to calculate the matrix-vector products in stage 2 (compare Sec. 3.2.2) efficiently and to reduce RAM requirements. In the simplest case of the FMM, mesh elements are clustered by a subdivision of the mesh's bounding cuboid. The edge length of the sub-cuboids is approximately $\sqrt[4]{N}$ times the average edge length of the mesh, and C is the number of clusters. Then, the kernels can be approximated by Eq. (6) and the system of equations reads as follows:

$$(9) \quad [\mathbf{N} + \mathbf{SDT}] \hat{\mathbf{x}} = \mathbf{b}.$$

where \mathbf{N} is the near-field matrix, \mathbf{S} describes the local expansion, \mathbf{D} is the translation matrix, and \mathbf{T} is the far-field signature matrix. The truncation order L is set to:

$$(10) \quad L = \max(8, 2.0 \cdot k \max_{\text{clusters}} r + 1.8 \log_{10}(2.0 \cdot k \max_{\text{clusters}} r + \pi)).$$

Given the truncation order, the integral over the unit-sphere (compare Eq. 6) is calculated by Gauss quadrature on the sphere with $K = 2L^2$ nodes [19, 20].

The actual generation of \mathbf{N} is similar to the generation of \mathbf{A} . However, the entries of \mathbf{N} are only calculated if the distance between the corresponding clusters is sufficiently large, and, thus, \mathbf{N} is sparse.

To calculate the entries of \mathbf{T} , the integral kernels for all combinations of Gauss quadrature nodes \mathbf{s}_k on the unit-sphere and field elements f_n in one cluster are integrated over Γ_{f_n} . The integrals over Γ_{f_n} are computed using standard Gauss quadrature and \mathbf{T} is a $CK \times N$ matrix.

\mathbf{D} consists of sums evaluated for Eq. (7). The spherical Hankel functions and the Legendre-polynomials are both calculated recursively. For the spherical Hankel function the modified Lentz method is used for efficient computation [26, 27]. \mathbf{D} is a $CK \times CK$ matrix.

The computation of \mathbf{S} is similar to the computation of \mathbf{T} . Instead of integrating over an element's surface, the integral over the unit-sphere is computed, which is done with the Gauss quadrature mentioned above [20].

The computational efficiency is further increased by applying the ML-FMM. For details refer to Ref. [19].

3.2.2 Stage 2: Solving

The conjugate gradient squared iterative solver [28] is used to calculate $\hat{\mathbf{x}}$. The stopping criterion is set to a relative residuum of $\frac{\text{resi}_i}{\text{reso}} < 10^{-10}$, where i is the iteration index. Incomplete LU-preconditioning is used to increase stability and convergence.

3.2.3 Stage 3: Post-Processing

When p and v on Γ are known, MESH2HRTF calculates the sound pressure at evaluation grid nodes \mathbf{x} by the use of the representation formula:

$$(11) \quad p(\mathbf{x}) = i\omega\rho \sum_{n=1}^N \int_{\Gamma_{f_n}} [H(\mathbf{x}, \mathbf{y})\phi_n - G(\mathbf{x}, \mathbf{y})v_n] d\mathbf{y} - i\omega\rho\phi_{inc}(\mathbf{x}), \quad \mathbf{x} \in \Omega_e.$$

3.3 Export

Two different file formats are used to store the results on the boundary surface and at the evaluation grid nodes.

The amplitude (in dB), and the phase (in degree) of the sound pressure and the velocity at collocations nodes and at evaluation grid nodes together with the geometric data are stored as *.vtk-files. For the visualization of results, these files can be directly loaded into PARAVIEW⁵, i.e., an open-source, cross-platform data analysis, and visualization application [29].

The complex spectra of the sound pressure on evaluation grid nodes are stored as *.sofa-files⁶ in the 'General-TF'-Convention [30]. SOFA is the AES-Standard for storing HRTFs [31]. The SOFA-API⁷ provides scripts for the conversion from the 'General-TF'-Convention to the 'SimpleFreeField-HRIR'-Convention and to calculate directional transfer functions.

⁵ Available from <http://www.paraview.org> (date last viewed: March 13, 2015)

⁶ More informations available from <http://www.sofaconventions.org> (date last viewed: March 13, 2015)

⁷ Available from <http://sourceforge.net/projects/sofacoustics/> (date last viewed: March 13, 2015)

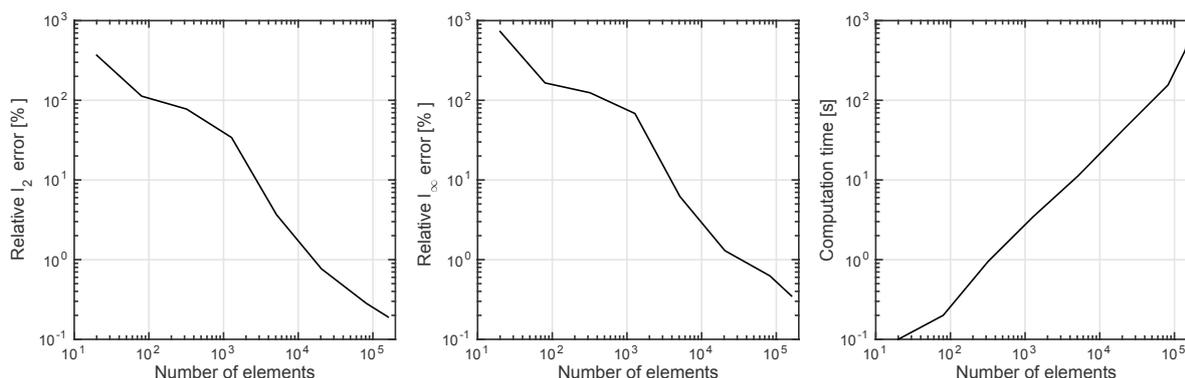


Figure 1: Relative error in the l_2 and the l_∞ sense and the computation time per frequency for computing \hat{x} .

4. Evaluation

4.1 Numerical Evaluation (sound-field scattered by a sphere)

A sphere was modeled as ico-sphere, i.e., a subdivided icosahedron, in BLENDER. The radius of the sphere was 10 cm and the sphere was placed at the center of the coordinate system. Two types of incident fields were evaluated. First, the sound-field of a point source placed at a radius of 20 cm directly to the left of the sphere was considered. Second, the sound-field of a plane wave propagating from the left to the right of the sphere was considered. Both, sound-soft (*Dirichlet*) and sound-hard (*Neumann*) boundary conditions were evaluated.

For the numerical evaluation, the number of elements in the mesh was systematically increased from 20 to 160 000 elements, the pressure was calculated at evaluation nodes in 1.2 m distance, the computation time was measured, and the relative error [32] in the l_2 and the l_∞ sense was calculated (see Fig. 1). The error was calculated relative to the analytical solution [33]. While the relative error was large for a mesh consisting of 20 elements, the relative error decreased rapidly below 1% for meshes with more than 80 000 elements. Because of the FMM, the computation time increased only by $N \log(N)$. Figure 2 shows the sound field in the horizontal plane for both boundary conditions and incident fields. The sound field obtained with the analytical solution for the sound-hard boundary condition is shown for visual comparison.

4.2 Perceptual Evaluation (HRTFs of human listeners)

The calculation of listener-specific HRTFs with Mesh2HRTF was extensively evaluated in Ref. [13, 14]. The results were evaluated by means of sound-localization models and in sound-localization experiments. The HRTFs calculated for meshes with more than 80 000 triangular elements yielded a sound-localization performance similar to the performance obtained with acoustically measured HRTFs.

5. Outlook

After finishing the process of publishing MESH2HRTF on SOURCEFORGE, it is planned to further improve the numerical core by implementing other iterative solvers, e.g., generalized minimal residual method (GMRES), by adding the Galerkin method and by parallelization of code blocks in OpenCL.

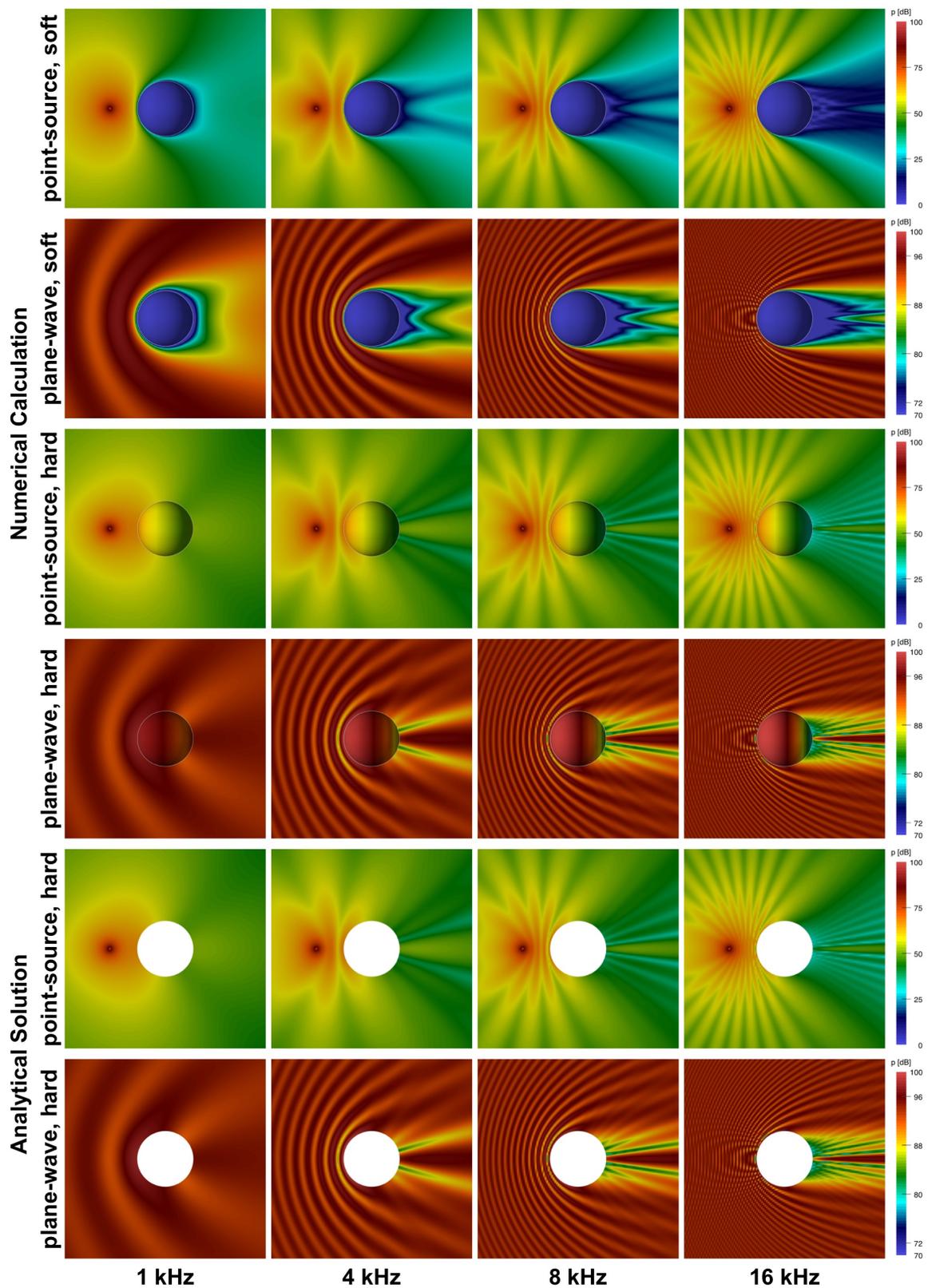


Figure 2: Numerically and analytically calculated sound pressure in the horizontal plane for a sound-soft sphere (point source and plane wave) and for a sound-hard sphere (point source and plane wave) at 1, 4, 8, and 16 kHz. The colorbars were limited for optimal visualization. Note that the analytical solution does not require the sound field on the surface of the sphere.

References

1. H. Møller, M. F. Sørensen, D. Hammershøi, and C. B. Jensen, "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.*, vol. 43, pp. 300–321, May 1995.
2. B. Xie, *Head-Related Transfer Function and Virtual Auditory Display*. J. Ross Publishing, 2 ed., 2013.
3. B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. rigid model calculation," *J. Acoust. Soc. Am.*, vol. 110, pp. 2440–2448, Nov. 2001.
4. T. Huttunen, E. T. Seppälä, O. Kirkeby, A. Kärkkäinen, and L. Kärkkäinen, "Simulation of the transfer function for a head-and-torso model over the entire audible frequency range," *J. Comp. Acous.*, vol. 15, pp. 429–448, Dec. 2007.
5. H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," *J. Acoust. Soc. Am.*, vol. 132, pp. 3832–3841, Dec. 2012.
6. L. Gaul, M. Kögl, and M. Wagner, *Boundary Element Methods for Engineers and Scientists: An Introductory Course with Advanced Topics*. Springer, 2003.
7. B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. II. impedance effects and comparisons to real measurements," *J. Acoust. Soc. Am.*, vol. 110, pp. 2449–2455, Nov. 2001.
8. Y. Kahana and P. A. Nelson, "Numerical modelling of the spatial acoustic response of the human pinna," *Journal of Sound and Vibration*, vol. 292, pp. 148–178, Apr. 2006.
9. Y. Kahana and P. A. Nelson, "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models," *Journal of Sound and Vibration*, vol. 300, pp. 552–579, Mar. 2007.
10. W. Kreuzer, P. Majdak, and Z. Chen, "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," *J. Acoust. Soc. Am.*, vol. 126, pp. 1280–1290, Sept. 2009.
11. N. A. Gumerov, A. E. O'Donovan, R. Duraiswami, and D. N. Zotkin, "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation," *J. Acoust. Soc. Am.*, vol. 127, pp. 370–386, Jan. 2010.
12. Y. Rui, G. Yu, B. Xie, and Y. Liu, "Calculation of individualized near-field head-related transfer function database using boundary element method," in *Proceedings of the 134th Convention of the Audio Engineering Society*, no. Convention Paper 8901, Audio Engineering Society, May 2013.
13. H. Ziegelwanger, A. Reichinger, and P. Majdak, "Calculation of listener-specific head-related transfer functions: Effect of mesh quality," in *POMA*, vol. 19, (Montreal, Canada), p. 050017, ASA, 2013.
14. H. Ziegelwanger, P. Majdak, and W. Kreuzer, "Requirements on the numerical calculation of listener-specific head-related transfer functions for accurate sound localization," *J. Acoust. Soc. Am.*, vol. submitted, 2015.
15. H. Ziegelwanger and P. Majdak, "Modeling the direction-continuous time-of-arrival in head-related transfer functions," *J. Acoust. Soc. Am.*, vol. 135, pp. 1278–1293, Mar. 2014.
16. H. Ziegelwanger, P. Majdak, and W. Kreuzer, "Non-uniform sampling of geometry for the numeric simulation of head-related transfer functions," in *Proceedings of the 21st International Congress of Sound and Vibration*, (Beijing, CN), 2014.
17. H. Ziegelwanger, P. Majdak, and W. Kreuzer, "Efficient numerical calculation of head-related transfer functions," in *Proceedings of the Forum Acusticum 2014*, (Krakow, Poland), p. 6, 2014.
18. A. J. Burton and G. F. Miller, "The application of integral equation methods to the numerical solution of some exterior boundary-value problems," *Proc. R. Soc. Lond. A*, vol. 323, pp. 201–210, June 1971.
19. Z.-S. Chen, H. Waubke, and W. Kreuzer, "A formulation of the fast multipole boundary element method for acoustic radiation and scattering from three-dimensional structures," *J. Comput. Acoust.*, vol. 16, pp. 303–320, 2008.
20. J. Rahola, "Diagonal forms of the translation operators in the fast multipole algorithm for scattering problems," *Bit Numer Math*, vol. 36, pp. 333–358, June 1996.
21. J. Möbius and L. Kobbelt, "OpenFlipper: An open source geometry processing and rendering framework," in *Proceedings of the 7th International Conference on Curves and Surfaces*, pp. 488–500, Springer, 2012.
22. M. Botsch, S. Steinberg, S. Bischoff, and L. Kobbelt, *OpenMesh - a generic and efficient polygon mesh data structure*. 2002.
23. I. H. Sloan and R. S. Womersley, "Extremal systems of points and numerical integration on the sphere," *Advances in Computational Mathematics*, vol. 21, pp. 107–125, July 2004.
24. S. E. Mousavi and N. Sukumar, "Generalized duffy transformation for integrating vertex singularities," *Computational Mechanics*, vol. 45, pp. 127–140, Jänner 2010.
25. G. Krishnasamy, L. W. Schmerr, T. J. Rudolph, and F. J. Rizzo, "Hypersingular boundary integral equations: Some applications in acoustic and elastic wave scattering," *J. Appl. Mech.*, vol. 57, pp. 404–414, June 1990.
26. K. Giebermann, *Schnelle Summationsverfahren zur numerischen Lösung von Integralgleichungen für Streuprobleme im R^3* . PhD thesis, Universität Karlsruhe, 1997.
27. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, Sept. 2007.
28. A. Meister, *Numerik linearer Gleichungssysteme*. Springer Spektrum, Nov. 2014.
29. U. Ayachit, *The ParaView Guide: A Parallel Visualization Application*. Kitware, Incorporated, Jan. 2015.
30. P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger, and M. Noisternig, "Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions," in *Proceedings of the 134th Convention of the Audio Engineering Society*, no. Convention Paper 8880, Audio Engineering Society, May 2013.
31. Audio Engineering Society, "AES standard for file exchange - spatial acoustic data file format," *AES69-2015*, 2015.
32. S. Marburg, "Six boundary elements per wavelength: Is that enough?," *J. Comput. Acoust.*, vol. 10, pp. 25–51, 2002.
33. L. L. Beranek and T. J. Mellow, *Acoustics: Sound Fields and Transducers*. Academic Press, Aug. 2012.

Chapter 3

Modeling the direction-continuous time-of-arrival in HRTFs

This work was published as:

Ziegelwanger, H., and Majdak, P. (2014). "Modeling the direction-continuous time-of-arrival in head-related transfer functions," *The Journal of the Acoustical Society of America*, **135**, 1278–1293. [doi:10.1121/1.4863196](https://doi.org/10.1121/1.4863196)

This study was collaborative research but based on my initial idea to model the time-of-arrival (TOA) by a geometric model, which I originally developed in my diploma thesis. With the help of the second author, I further developed the TOA model. Together, we developed the outlier detection and designed the evaluation of the TOA model. For the evaluation, I numerically calculated HRTFs for various geometrical objects and analyzed the TOA model parameters when applied to HRTFs from three databases. The writing of the manuscript was done in close collaboration between the authors where I learned how to write a scientific journal article. While the second author was responsible for Fig. 8, I generated all other figures.

Co-author:

Modeling the direction-continuous time-of-arrival in head-related transfer functions

Harald Ziegelwanger^{a)} and Piotr Majdak

Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria

(Received 4 April 2013; revised 24 December 2013; accepted 3 January 2014)

Head-related transfer functions (HRTFs) describe the filtering of the incoming sound by the torso, head, and pinna. As a consequence of the propagation path from the source to the ear, each HRTF contains a direction-dependent, broadband time-of-arrival (TOA). TOAs are usually estimated independently for each direction from HRTFs, a method prone to artifacts and limited by the spatial sampling. In this study, a continuous-direction TOA model combined with an outlier-removal algorithm is proposed. The model is based on a simplified geometric representation of the listener, and his/her arbitrary position within the HRTF measurement. The outlier-removal procedure uses the extreme studentized deviation test to remove implausible TOAs. The model was evaluated for numerically calculated HRTFs of sphere, torso, and pinna under various conditions. The accuracy of estimated parameters was within the resolution given by the sampling rate. Applied to acoustically measured HRTFs of 172 listeners, the estimated parameters were consistent with realistic listener geometry. The outlier removal further improved the goodness-of-fit, particularly for some problematic fits. The comparison with a simpler model that fixed the listener position to the center of the measurement geometry showed a clear advantage of listener position as an additional free model parameter. © 2014 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4863196>]

PACS number(s): 43.60.Uv, 43.66.Yw, 43.60.Jn, 43.66.Pn [ZHM]

Pages: 1278–1293

I. INTRODUCTION

Head-related transfer functions (HRTFs) are direction-dependent filters describing the acoustic filtering of the head, torso, and pinna (Blauert, 1997; Mehrgardt and Mellert, 1977; Møller *et al.*, 1995). They provide cues for the localization and externalization of virtual sound sources presented via headphones, a procedure known as binaural reproduction technique (Wightman and Kistler, 1989). With a complete set of room-specific HRTFs, arbitrary combinations of virtual sound sources can be presented via so called virtual auditory displays (Xie, 2013).

Binaural pairs of HRTFs provide interaural time differences (ITDs) which are important for the lateralization of sound-sources in the horizontal plane (Strutt, 1894; Macpherson and Middlebrooks, 2002). While the ITD is a relative quantity and thus sufficient for the description of an acoustic scene with a single sound source, the inter-source time difference, i.e., the delay between individual sound sources is required for an acoustically correct representation of multiple sound sources in a scene. This inter-source time difference can only be derived when the absolute delay arising from the acoustic wave propagation path from each source to a listener's ear is available. Those delays can be described by the directional time-of-arrival (TOA), i.e., a delay in an HRTF as a function of the sound direction. Note that for a given direction, the difference between the left-ear TOA and the right-ear TOA corresponds to the broadband¹ ITD.

The TOA can be derived from the spectral phase of an HRTF by separating the spectral phase in three components:

Minimum phase, frequency-independent TOA, and excess phase. While the minimum phase can be calculated with the Hilbert transform (Oppenheim *et al.*, 1999), a robust estimation of the TOA from an impulse response is not trivial (Defrance *et al.*, 2008). In the research field of binaural hearing, most of the TOA estimation algorithms consider single directions only, mainly focusing on ITD estimation. Because of the close relation between the TOA and the ITD, some of the ITD estimation methods also implicitly estimate the TOA for both ears. These methods can be classified in frequency-domain (Huopaniemi and Smith, 1999; Jot *et al.*, 1995), time-domain (Jot *et al.*, 1995; Møller *et al.*, 1995), and other methods (Nam *et al.*, 2008; Wightman and Kistler, 2005). All these methods are susceptible to measurement noise, and even though low-pass or high-pass filtering have been used to improve the estimation stability (Algazi *et al.*, 2001a), the methods produce artifacts, which correction is not trivial (Busson *et al.*, 2005). The artifacts can be found by following the TOA along the direction but can not be considered by an estimator processing each sound direction separately. The detection and correction is thus, usually done *post hoc* in a manual procedure.

In other research fields, sophisticated methods like particle filtering considering spatial evolution of TOA (Michalopoulou and Jain, 2012) have been developed to overcome these problems. Applied to the field of HRTFs, in such a *joint-directional* timing model, the timing information (TOA or ITD) of the sound propagation paths around the human head is modeled as a continuous function of direction. A simple example for such a model is the ITD model for the horizontal plane (Kuhn, 1977), which describes ITD by two model parameters: The radius of the circle and the orientation of the interaural axis. This model assumes that

^{a)}Author to whom correspondence should be addressed. Electronic mail: harald.ziegelwanger@oeaw.ac.at

ITD can be correctly described by a function of the lateral angle in the horizontal plane and has been extended to a two-dimensional model (Savioja *et al.*, 1999). In order to estimate the parameters of such a model for a specific listener, the timing estimations from HRTFs measured for different directions are used to fit a continuous function of direction. In such a model, potentially incorrect timing estimations do not *define* the timing for the corresponding direction, but they, together with the correct estimations, *contribute* to a model as a continuous function of direction. Wang *et al.* (2009) proposed to fit a polynomial function with truncated order to model ITD as a two-dimensional function of direction. Even though this model could be a starting point for a TOA model, the parameters of the polynomial function have a very weak relation to the geometry of the measurement setup and physical interpretation of such model parameters seems to be rather difficult. Thus, in this study, we propose and investigate the use of a continuous-direction model of the TOA with a close relation to the geometry of the HRTF measurement setup.

II. GENERAL METHODS

A. Coordinate systems

Two coordinate systems are used for the directional representation of HRTF data. First, we use the geodesic coordinate system (Fig. 1, left panel), in which the direction of a sound source is described by the azimuth ϕ and elevation θ angle. A horizontal plane is selected by an elevation angle. The horizontal plane selected by $\theta = 0^\circ$ is called the interaural horizontal plane. Second, we use the interaural-polar coordinate system, which is the geodesic coordinate system rotated such that the polar axis becomes equivalent with the interaural axis (Fig. 1, right panel). In the interaural-polar coordinate system, a sagittal plane is selected by the lateral angle Φ and the direction of a source within a sagittal plane is described by the polar angle Θ . This coordinate system more distinctly addresses the human sound-localization mechanisms because the lateral direction is mostly associated with binaural cues and polar direction is mostly associated with spectral cues (Morimoto and Aokata, 1984).

B. HRTF material

In order to evaluate the proposed models, parameters resulting from the models were compared to the parameters

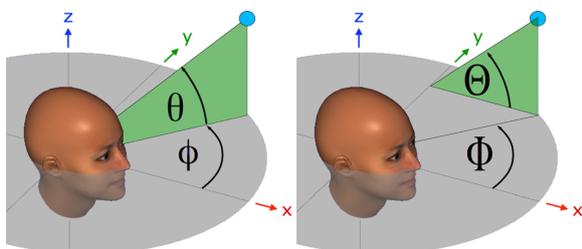


FIG. 1. (Color online) Coordinate systems. Geodesic system (left panel) described by the azimuth ϕ and the elevation θ angles. Horizontal-polar system (right panel) described by the lateral Φ and the polar Θ angles.

actually used in the HRTF measurements. In order to provide well-defined actual parameters, we used numerically calculated HRTFs, for which the simulated measurement setup can be perfectly constructed. Further, the model was applied to acoustically measured HRTFs of human listeners. The exact parameters actually used in HRTF measurements, such as position, rotation, and size of the listener's head, can not be perfectly reconstructed and thus are not used for the evaluation.

1. Numerical calculation

HRTFs of three different objects were numerically calculated. The first object, Sphere, was a rigid sphere with two virtual microphones arbitrarily positioned on the surface. The second object, sphere-and-torso (SAT), was the object Sphere combined with a simplified torso placed below the sphere. The third object was sphere-torso-pinna (STP), which was the SAT with a stitched pinna on the sphere, see Fig. 2. Even though the contribution of the pinna was expected to be rather small, the STP is considered as a relevant link to the actual morphology of a human listener.

The sphere and the torso were constructed based on icosphere objects in Blender.² The torso was an ellipsoid with axes of 200 mm \times 520 mm \times 180 mm (width \times depth \times height) with its center positioned 190 mm below the center of the sphere. The torso and sphere were linked with a neck, which was modeled as a cylinder with a radius of 50 mm. The pinna was created in MakeHuman³ with the application's default parameters. All objects were meshed to an average edge length of 3 mm. Top and side view on the STP object is shown in Fig. 2.

For all objects, the sound pressure in the simulated free-field was calculated with the Burton-Miller collocation boundary-element method coupled with the multi-level fast-multipole method (for more details see Kreuzer *et al.*, 2009). Complex spectra were calculated in the range of 0.2 to 20 kHz with a frequency resolution of 0.2 kHz. The principle of reciprocity was used to decrease the computational costs in the calculation. HRTFs were calculated for 1550 directions at a 3-m distance. The directions were in the elevation between -30° to 80° in steps of 5° . The azimuthal range was 360° in steps of 2.5° within the interaural horizontal plane

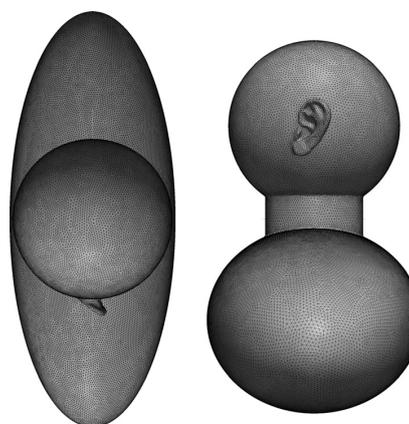


FIG. 2. Top and side view of the object STP.

and larger steps within other horizontal planes in order to obtain an approximately constant directional resolution. Finally, the head-related impulse responses (HRIRs) were obtained by applying the inverse Fourier transform on the complex spectra and resampling to 48 kHz. The final duration of the HRIRs was 5 ms.

The actual parameters, for which the HRTFs were calculated are described in the corresponding sections. In total, HRTFs of 98 objects were calculated.⁴

2. Acoustic measurements

Acoustically measured HRTFs of in total 172 human listeners from three HRTF databases were considered.

The database ARI⁵ contains HRTFs measured in a semi-anechoic sound chamber with the blocked-ear-canal method (Møller *et al.*, 1995) for 1550 directions at a distance of 1.2 m. The horizontal and vertical range was 360° and −30° to 80°, respectively, with the horizontal and vertical resolution of 2.5° and 5°, respectively (for more details see Majdak *et al.*, 2010). In our study, the HRTFs of 77 listeners from NH2 to NH132 were used (not all IDs are populated between NH2 and NH132). In text, this database is referred to as DB1.

The database CIPIC⁶ contains HRTFs measured in an acoustic facility with the blocked-ear-canal method for 45 listeners and 1250 directions at a distance of 1 m. The horizontal and vertical range of directions was 360° and −80° to 80°, respectively, with the horizontal and vertical resolution of 5.625° and 5°, respectively (for more details see Algazi *et al.*, 2001b). In text, this database is referred to as DB2.

The database LISTEN⁷ contains 51 HRTF sets measured in an anechoic sound chamber with the blocked-ear-canal method for 187 directions at a distance of 1.95 m. The horizontal and vertical range was 360° and −45° to 90°, respectively, with the horizontal and vertical resolution of 15° and 15°, respectively. HRTFs of subject #34 have been excluded from our study because of some directions containing invalid HRIRs in terms of the total energy of 30 dB below that for the other directions. In text, this database is referred to as DB3.

III. TOA ESTIMATORS

In this section, we review the relevant TOA estimators and investigate their results when applied to HRTFs.

A. General

The estimation of TOA relies on the separation of the minimum phase, frequency-independent TOA, and excess phase in an HRTF. While most of the existing methods focus on the estimation of ITD, some of them implicitly estimate the TOA. The methods can be classified as time-domain, frequency-domain, and mixed methods.

Frequency-domain methods rely on the phase spectrum. In order to estimate the TOA, a line is fit to the phase spectrum and the slope of the line is defined as the TOA (Huopaniemi and Smith, 1999; Jot *et al.*, 1995). The TOA has also been derived from the group delay, i.e., the frequency derivative of the phase, by averaging the group delay

over the frequency range of 1 to 5 kHz (Jot *et al.*, 1995), or the frequency range of 0.5 to 2 kHz (Huopaniemi and Smith, 1999). These frequency-domain methods suffer from the phase wrap around 2π and the corresponding difficulties in the unwrap algorithm when a change of $-\pi$ or $+\pi$ cannot be distinguished (Tribolet, 1977). The unwrap errors directly affect the TOA estimation. Even though the group delay can be directly calculated from the complex spectrum, it still might yield noisy results (Raykar *et al.*, 2005).

Time-domain methods do not suffer from the phase wrap. The simplest time-domain method for the TOA estimation is the amplitude threshold, in which TOA is the time where a (pre-processed) HRIR reaches a threshold for the first time. In particular, the following combinations of pre-processing and thresholds have been used: The maximum (or 20 dB below the maximum) of the absolute-valued HRIR (Møller *et al.*, 1995), its envelope (Busson *et al.*, 2005), its low-passed (Algazi *et al.*, 2001a), or its high-passed versions. In a similar time-domain method, the centroid of the HRIR has been used to calculate the ITD and therefore might be interpreted as TOA (Jeppesen and Møller, 2005; Jot *et al.*, 1995). The correspondence of those methods to the frequency-domain methods is not simple. As a combination of time- and frequency method, zero-crossing method has been used to estimate the group delay in the time domain (Lee *et al.*, 2009), but it yielded noisy results for realistic signals.

Based on the generalized cross-correlation method (Knapp and Carter, 1976), an interesting TOA estimator has been proposed (Wightman and Kistler, 2005; Nam *et al.*, 2008), where the HRIR h and its minimum-phase version h_{\min} were cross-correlated and the resulting maximum was evaluated,

$$\hat{\tau} = \operatorname{argmax}_k \left| \sum_{n=-\infty}^{\infty} h[n]h_{\min}[n+k] \right|, \quad (1)$$

with $\hat{\tau}$ as the estimated TOA and n as the sample index corresponding to the discrete time. Note that h_{\min} can be easily calculated from the causal part of an HRTF's complex cepstrum (Oppenheim *et al.*, 1999).

This method has some interesting properties. First, it might be less vulnerable to phase artifacts in spectral notches because the amplitude spectrum of the cross-correlation is the squared amplitude spectrum of the HRTF, and thus, frequency regions with little energy do not contribute much to the TOA estimation. Second, if an HRTF is a "minimum-phase plus delay" system, which seems to be approximately valid for at least some directions of the incoming sound (Kulkarni *et al.*, 1999; Mehrgardt and Mellert, 1977) then the excess phase is approximately zero. Consequently, the cross-correlation in Eq. (1) reduces to a delayed pulse in the time domain, whose delay can be easily estimated and corresponds *exactly* to the TOA. Thus, this method works well for directions for which HRTFs are approximately "minimum-phase plus delay" systems. Interestingly, the transfer function of the pressure at the surface of a sphere is minimum phase (Constan and Hartmann, 2003, p. 1007 and Fig. 6).

However, HRTFs of a human can not be minimum-phase systems because of the reflections at the torso and the pinna. Pinna contributes in higher frequencies only and torso contributes in lower frequencies, with stronger contributions from the ipsilateral side (Blauert, 1997). While the impact of these contributions on ITDs is frequency dependent and rather small (Kuhn, 1977, Figs. 11, 12, and 13), their impact on the excess phase in HRTFs seems to be not clarified yet.

Thus, in our study, we do not rely on the assumption that HRTFs are minimum-phase systems. We rather apply different TOA estimators on HRTFs and investigate their compatibility with our model.

B. Application

In order to estimate the impact of the TOA estimators, results from various TOA estimators were investigated when applied on different HRTFs, with a particular focus on the impact of torso, pinna, and the acoustic measurement noise. To this end, numerically calculated HRTFs represented the noise-free condition in which HRTFs of SAT and STP were used to demonstrate the impact of torso and pinna in comparison to HRTFs of a sphere. Acoustically measured HRTFs of an exemplary listener (NH89, DB1) were used to investigate the usefulness of TOA estimators under more realistic conditions.

The following TOA estimators were considered: (1) TOA at the maximum of the absolute HRIR (MAX; Møller *et al.*, 1995); (2) TOA at the centroid of the HRIR (CTD; Jeppesen and Møller, 2005); (3) TOA as the average group delay of the HRTF, calculated from the negative frequency derivative of the phase, and averaged between 1 and 5 kHz (AGD; Jot *et al.*, 1995); and (4) TOA at the maximum of the cross-correlation between the HRIR and its minimum phase version (MCM; Nam *et al.*, 2008) according to Eq. (1).

Figure 3 shows estimated TOAs for the interaural horizontal plane with one TOA estimator for each panel. The torso does not seem to have affected the estimated TOAs much. For example, for the MCM estimator, the difference between the TOAs of SAT and that of the Sphere was $0.4 \mu\text{s}$

(average over all directions). Compared to torso, the pinna seems to have an effect on the estimated TOA. For the MCM estimator, the average difference between the TOAs of SAT and that of STP was $-12.0 \mu\text{s}$. It seems like the combination of the pinna and torso affected the TOA estimation substantially, but the effect of the torso alone was negligible.

TOAs obtained for the acoustically measured HRTFs seem to follow a different pattern with artifacts for particular directions. These findings support others' observations that TOA estimation from acoustically measured HRTFs is a difficult task. Especially the artifacts might yield larger errors in modeling and might require a special handling.

Even though the comparison between the different TOA estimators was only used to illustrate their results, the differences are evident. Especially, the MAX estimator seems to produce many artifacts. Also, the CTD estimator seems to result in systematically larger TOAs than the MCM estimator. Nevertheless, at this stage, we are not able to reliably judge the compatibility of the TOA estimators with the proposed model, and thus, in the following section, we consider all of them in the evaluation of the model.

IV. ON-AXIS TOA MODEL

In this section, the directional TOA model is derived. The model aims at describing simplified geometric properties of the listener and the HRTF measurement setup. The model is fit to TOAs estimated by one of the previously described TOA estimators, and the results are evaluated.

A. Model derivation

The listener's head is modeled as a rigid sphere that is placed exactly in the center of the measurement system. Thus, the sphere center and the measurement center are assumed to be coincident. The model further assumes a stationary position of listener and the loudspeakers. Figure 4 illustrates the model with the arbitrary position $\vec{e} = [\phi_e, \theta_e]$ of the ear on the sphere. The sound propagation path

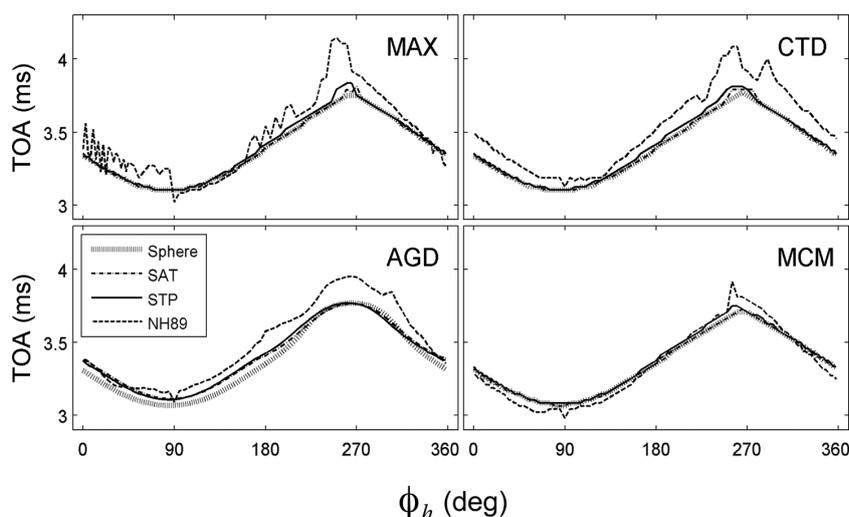


FIG. 3. TOAs resulting from TOA estimators in the horizontal interaural plane applied on calculated HRTFs of the objects Sphere, SAT, and STP, as well as on measured HRTFs of an exemplary listener (NH89, DB1).

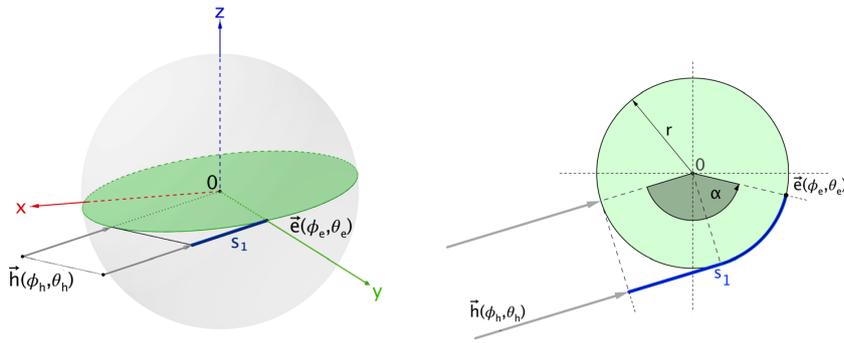


FIG. 4. (Color online) On-axis model. Left panel: Sphere representing the head placed in the origin of the coordinate system. The circle represents the great circle defined by the direction of the incoming wave (\vec{h}), the center of the sphere (0), and the position of the ear (\vec{e}). Right panel: Projected view on the great circle showing the propagation path s_1 defining the TOA.

between a sound source and the listener's ear is modeled corresponding to the sound propagation path from a loudspeaker to the in-ear-canal microphone in the HRTF measurement. A plane-wave propagation is assumed, thus, the curvature of the incident sound waves generated by the sound sources are assumed to be small in relation to the entire sound propagation paths. TOA is modeled as the time the sound requires to travel along the propagation path for a given direction.

For a given direction $\vec{h} = [\phi_h, \theta_h]$ of a sound source, the TOA is split in a direction-independent and direction-dependent part. The direction-independent TOA $\bar{\tau}$ is the propagation path between a loudspeaker and the closest point on the sphere. For the direction-dependent TOA and its corresponding sound propagation path, the cases for the ipsilateral and the contralateral sound sources are considered separately. The distinction is based on the angular distance α , which is

$$\alpha = \arccos(\vec{e} \cdot \vec{h}). \quad (2)$$

For the ipsilateral sound sources $\alpha \leq \pi/2$, the direction-dependent propagation path s_1 is

$$s_1 = r[1 - \sin \theta_e \sin \theta_h - \cos \theta_e \cos \theta_h \cos(\phi_e - \phi_h)] \quad (3)$$

for $\alpha \leq \frac{\pi}{2}$.

For the contralateral sound sources $\alpha > \pi/2$, the direction-dependent propagation path additionally considers the scattering path around the sphere and is

$$s_1 = r + r \arccos[\sin \theta_e \sin \theta_h + \cos \theta_e \cos \theta_h \cos(\phi_e - \phi_h)] - \frac{r\pi}{2} \quad \text{for } \alpha > \frac{\pi}{2}.$$

With the speed of sound c , the TOA is then

$$\tilde{\tau} = \frac{s_1}{c} + \bar{\tau}. \quad (5)$$

In order to obtain the model parameters $\bar{\tau}$, r , ϕ_e , and θ_e , Eq. (5) is fit to the set of estimated TOAs $\hat{\tau}$. The fit is done by minimizing the squared error between $\tilde{\tau}$ and $\hat{\tau}$ using an iterative trust-region-reflective algorithm implemented as *lsqcurvefit*.⁸ This algorithm is a subspace trust-region method and is based on the interior-reflective Newton method (Coleman and Li, 1996). It implements a robust and efficient

globalization strategy for computing a local minimum of a possibly non-convex and once continuously differentiable function. In each iteration, the algorithm approximately solves a linear system using the method of preconditioned conjugate gradients, until the termination tolerance is reached. In our study, the termination tolerance was set to 10^{-6} . For a general review of nonlinear least-squares methods, see Dennis (1977).

The algorithm converges at a local minimum, which distance to the global minimum depends on the initial values. Thus, providing reasonable initial values is important. For example when fitting the model to HRTFs of human, the initial values should correspond to average human parameters. Further, trivial values like $r=0$ must be avoided. In this study, the initial values were chosen in each evaluation and/or application separately depending on the origin of the data.

Finally, the algorithm requires constrains. For the on-axis model, the constrains were always the same, namely, ear position constrained to $\phi_e = \pm 180^\circ$ and $\theta_e = \pm 90^\circ$ (effectively unconstrained) and the radius constrained to ± 6 cm of the radius initial value.

Note that the model is fit to estimated TOAs for each ear separately.

B. Evaluation

In order to evaluate the goodness of the model fit, we used the adjusted norm of residuals (ANR), i.e., the square root of the adjusted sum of squares of residuals (Bajorski, 2011). The adjustment happens by dividing the sum by the number of residuals, which makes the norm less dependent on the number of HRTF directions. ANR varies from 0 to infinity—the smaller the norm the better the fit. It provides an idea about the average (in terms of root-mean-square) size of the residuals per fit.

1. Effect of parameter variation

The on-axis TOA model was evaluated addressing the following: (1) general ability of the model to describe TOA; (2) ability to describe TOAs estimated from HRTFs measured under conditions represented by spheres of different sizes and different microphone positions; (3) effect of initial values on the goodness of fit. To this end, HRTFs were calculated for the object Sphere under conditions resulting from

TABLE I. ANRs and parameter errors (average ± 1 standard deviation) resulting from fitting the on-axis model to TOAs estimated from HRTFs of the object Sphere (135 conditions, compare Sec. IV B 1). EST: TOA estimator. Parameter errors: Differences between the estimated and actual parameters. The averages and standard deviations are across the tested conditions.

EST	r error (mm)	ϕ_e error ($^\circ$)	θ_e error ($^\circ$)	ANR (μ s)
MAX	2.8 ± 0.2	1.1 ± 0.0	0.8 ± 0.0	6.5 ± 0.1
CTD	3.2 ± 0.2	1.2 ± 0.0	0.9 ± 0.1	6.9 ± 0.1
AGD	15.4 ± 0.9	1.9 ± 0.0	2.5 ± 0.0	23.6 ± 0.6
MCM	1.5 ± 0.2	1.1 ± 0.0	0.6 ± 0.0	6.1 ± 0.0

the following parameter combinations: r was 77.5, 87.5, and 97.5 mm, ϕ_e was -40° to $+40^\circ$ in steps of 10° , θ_e was -10° to $+10^\circ$ in steps of 5° . Each combination of the actual parameters is referred to as a condition. The position of the left and right ear microphones was varied independently, however, only the most relevant combinations of parameters were tested. Note that $r = 87.5$ mm addresses the average head radius of human population (Hartley and Fry, 1921). Four TOA estimators (MAX, CTD, AGD, and MCM) were used to estimate TOAs from each HRTF set. For the fits, the initial values were $r = 87.5$ mm, $\phi_e = \pm 90^\circ$, and $\theta_e = 0^\circ$. Initial values were identical in all conditions.

Table I shows the ANRs and the parameter errors, i.e., differences between the actual and estimated parameters, averaged over all conditions for each of the TOA estimators. Recall that the smaller ANR, the smaller the difference between the modeled and estimated TOAs, and thus, the better the model represents the estimated TOAs. Also, the smaller the parameter error, the more accurate the model estimated the actual parameter of the geometry.

For the MAX, CTD, and MCM estimators, the ANRs were well below the sampling interval of 20.8μ s. The sphere radius errors were within the spatial resolution of the propagation paths (7.1 mm at $c = 340$ m/s and sampling rate of 48 kHz). The ear position errors were within the range of 1.2° . For the MCM estimator, Fig. 5 shows the estimated model parameters for each condition.

For the AGD estimator, the ANR was three times larger than for the other estimators. Also the parameter errors were larger than for the other estimators. The estimated radii were on average 15.5 mm larger than the actual radii used in the HRTF simulations.

The fits seem to be robust to the initial values: Identical initial values were used in all conditions, and still the estimated model parameters converged near the actual values.

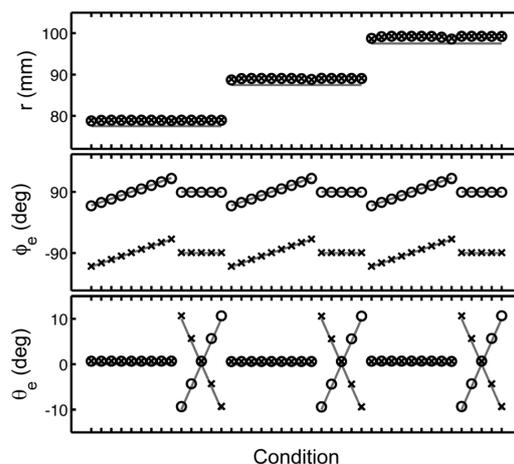


FIG. 5. Estimated on-axis model parameters from HRTFs calculated for the centered object Sphere. Condition: combination of actual parameters used in HRTF simulations. Circle and cross: Parameters estimated for the left and right ears, respectively. Gray lines: Actual parameters.

2. Effect of sphere, torso, and pinna

The proposed model approximates the listener's head as a sphere. When applied to HRTFs of human listeners, torso and pinna might affect the model results. In order to estimate the impact of torso and pinna, the on-axis model was fit to TOAs estimated for calculated HRTFs of the objects Sphere, SAT, and STP. For the HRTF calculations, the sphere radius was 87.5 mm and the virtual microphones were placed at typical ear positions, i.e., $\phi_e = \pm 85^\circ$ and $\theta_e = -10^\circ$. Four TOA estimators (MAX, CTD, AGD, and MCM) were used to estimate TOAs, then the on-axis model was fit to the estimated TOA sets. The initial values were $r = 90$ mm, $\phi_e = \pm 90^\circ$, and $\theta_e = 0^\circ$, i.e., slightly different than the actual parameters of the HRTF calculations.

Table II shows ANRs and model parameters for the four TOA estimators. Generally, the results seem to depend on both TOA estimator and evaluated object. For MAX, CTD, and MCM estimators, the ANRs were below 10.4μ s, which is only a fraction of the sampling interval of 20.8μ s. The estimated ear positions had errors of 3° or smaller, the estimated radii had errors of 7.7 mm or smaller. For the AGD estimator, the ANR increased up to 54.1μ s, the estimated ear position and radius had errors as large as 4.3° and 17.5 mm, respectively. Such large errors indicate that the AGD estimator is not compatible with the proposed model.

TABLE II. Parameters and ANRs resulting from fitting the on-axis model to TOAs estimated from HRTFs of the objects Sphere, SAT, and STP. Actual parameters: $r = 87.5$ mm, $\phi_e = 85^\circ$, and $\theta_e = -10^\circ$.

	r (mm)				ϕ_e ($^\circ$)				θ_e ($^\circ$)				ANR (μ s)			
	MAX	CTD	AGD	MCM	MAX	CTD	AGD	MCM	MAX	CTD	AGD	MCM	MAX	CTD	AGD	MCM
Sphere	91.4	91.9	105.0	90.2	84.9	84.9	84.3	84.9	-9.6	-9.7	-8.3	-9.9	6.1	6.5	19.8	6.3
SAT	92.1	92.9	98.8	90.4	84.9	84.9	85.1	84.9	-9.0	-8.9	-7.2	-9.6	8.0	9.8	54.1	6.6
STP	95.4	95.4	100.7	93.2	82.9	82.8	81.5	83.2	-9.8	-9.6	-5.7	-9.4	9.4	10.4	52.8	8.8

For the remaining TOA estimators, the ANRs were fractions of the sampling interval, even when the torso and pinna were considered in HRTF simulations. This indicates that the proposed model was generally able to describe the TOAs even in the presence of torso and pinna. The estimated radii increased, however, when the torso and even more when the pinna were considered in HRTF simulations. For MAX, CTD, and MCM estimators, the radius error increased from 3.9, 4.5, and 2.7 mm (Sphere), respectively, to 7.9, 7.9, and 5.7 mm (STP), respectively. It seems like the pinna had the most effect on the estimated radius—from Sphere to SAT, the increase in the radius error averaged over the three estimators was 0.6 mm, from SAT to STP, it was 2.9 mm. This indicates that pinna, not the torso, had the most prominent effect on the estimated radii. This effect might be related to the longer propagation paths around the pinna, an aspect not considered in the proposed model.

The combination of the on-axis model with the MCM estimator showed the most accurate parameter estimations, even in the presence of pinna and torso. For STP, the radius error was 5.7 mm which is within the spatial resolution of the propagation paths of 7.1 mm. The estimated ear position had an error of 1.8° . We thus, consider the MCM estimator as an estimator best compatible with the proposed model and focus on this estimator in the following sections.

C. Application

The model is intended to be applied on acoustically measured HRTFs of human listeners. Thus, in this section, we analyze the results of fitting the model to TOAs obtained from HRTFs of human listeners (described in Sec. II B 2). For each listener and ear, TOAs were estimated using the MCM estimator and the model was fit. The initial values were $r = 87.5$ mm, $\phi_e = \pm 90^\circ$, and $\theta_e = 0^\circ$.

Figure 6 shows the estimated model parameters as histograms over all listeners. Each histogram was fit to a Gaussian function in order to estimate the average and the standard deviation of the distributions. The statistics of the model parameters and ANRs are shown in Table III. The ANRs are positive values and can be approximated by the normal distribution only when their distribution has a small variance. Thus, for ANRs, medians were calculated. Further, the interaural radius differences (IRDs), i.e., the listener-specific differences between the radii estimated for the left and the right ears were calculated. The average IRD was

12.5 ± 22.3 mm (average \pm standard deviation), with individual IRDs ranging from -57.2 to 80.8 mm, see Fig. 7 (top panel). For the three databases DB1, DB2, and DB3, the average IRD was 11.0, 4.8, and 21.6 mm, respectively. The *unsigned* IRD was 19.8 ± 16.2 mm. For the three databases, the average of unsigned IRDs was 21.9, 7.2, and 27.7 mm, respectively.

D. Discussion

The on-axis model describes an HRTF measurement with a listener placed in the center of the coordinate system. The evaluation performed on the calculated HRTFs showed a good correspondence between the actual parameters used for HRTF simulations and the parameters estimated by the on-axis model. The model relies on a TOA estimator, and the model using the MCM estimator showed the best correspondence between the actual and estimated parameters. For the MCM estimator, the ANRs were in the range of few microseconds, which is a small fraction of the sampling interval, indicating a good fit of the model to the data. The differences between the estimated and the actual ear positions were smaller than 1.6° , which is smaller than the minimum audible angle being in the range of 2° (Mills, 1958). The differences between the estimated and the actual radius were smaller than 5.6 mm, which is within the resolution of the propagation paths given by the sampling interval and the speed of sound.

The model was then applied to acoustically measured HRTFs of 172 listeners. Note that for acoustically measured HRTFs, an evaluation in terms of comparing the estimated and actual parameters can not take place. The estimated parameters can be, however, checked for plausibility. For example, the ear position or the sphere radius should correspond to those usually found in humans. Further, the model independently fit to TOAs obtained for the two ears of the same listener should yield in similar estimated radii. In other words, an IRD substantially deviating from zero should raise evidence for implausible results.

Interestingly, the check of plausibility showed mixed results. The binaural average of the estimated radius was with 86.95 mm close to the average head radius of adult humans of 87.5 mm (Hartley and Fry, 1921). Also the estimated average ear positions, nearly most lateral (ϕ_e of $\pm 88^\circ$) and a little below the head center (θ_e of -7°) seem to be close to those usually found in humans

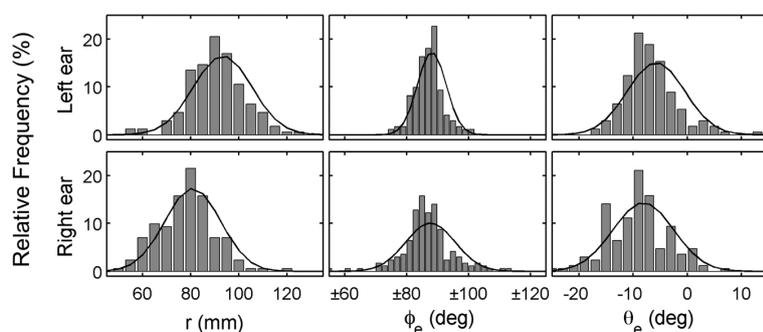


FIG. 6. Relative frequencies (%) of estimated on-axis model parameters from left-ear (top row) and right-ear (bottom row) HRTFs of human listeners. Lines: normal distribution fitted to the data. ϕ_e : Positive and negative values correspond to the left and the right ears, respectively.

TABLE III. Parameters and ANRs (average ± 1 standard deviation) resulting from fitting the on-axis model to TOAs estimated from acoustically measured HRTFs of human listeners. L: Left ear, R: Right ear. All: Results for all listeners. NH89: Results for a single listener (NH89, DB1).

	Ear	r (mm)	ϕ_e ($^\circ$)	θ_e ($^\circ$)	ANR (μ s)
All	L	93.2 ± 12.1	88.1 ± 4.6	-5.7 ± 5.3	40.9 ± 47.1
	R	80.7 ± 11.5	-87.7 ± 8.0	-8.0 ± 5.6	38.4 ± 41.4
DB1	L	88.9 ± 6.2	86.4 ± 3.4	-6.7 ± 3.0	49.4 ± 17.4
	R	84.1 ± 5.1	-86.2 ± 4.1	-8.1 ± 3.4	47.9 ± 16.3
DB2	L	92.9 ± 12.5	87.6 ± 4.3	-6.7 ± 3.1	27.1 ± 8.3
	R	81.9 ± 12.4	-87.4 ± 7.6	-8.6 ± 5.1	29.2 ± 7.8
DB3	L	97.7 ± 14.0	90.4 ± 5.2	-3.6 ± 8.2	54.6 ± 82.5
	R	76.1 ± 12.9	-89.5 ± 10.5	-7.5 ± 7.6	44.0 ± 73.4
NH89	L	107.3	82.3	-7.8	29.9
	R	68.5	-72.6	-12.6	38.7

(Algazi *et al.*, 2001b; Burkhard and Sachs, 1975). There are, however, two serious concerns about the results.

First, the ANRs (binaural average of 23.6μ s) were much larger than those obtained for calculated HRTFs (binaural average for the MCM estimator of 6.1μ s). The larger ANRs might result from artifacts in the TOA estimations, which underlying reason might be noise and inaccuracies involved in the acoustic measurements. In order to reduce the larger ANRs, in the next section, we propose to handle these artifacts before fitting the model.

Second, the estimated radii substantially differed between the two ears of each listener. For one listener, the IRD was even as large as 80.8 mm. Such large discrepancy raises evidence that the on-axis model might be a too simple representation of HRTF measurement geometry. This issue seems to be independent from the larger ANRs, as supported by the rather low correlation coefficients for the correlation between the IRD and the ANR of 0.22 and -0.05 for the left and right ears, respectively. Thus, in the two next sections,

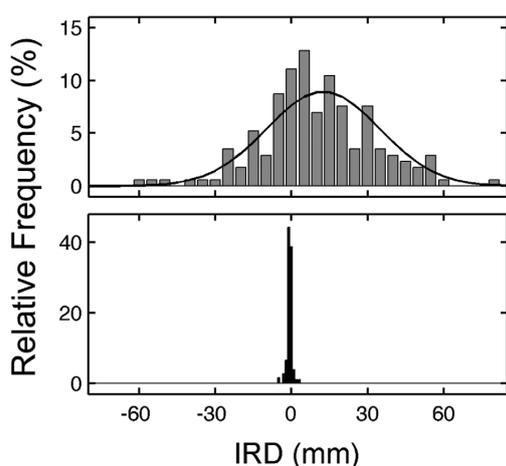


FIG. 7. Estimated on-axis model IRDs (top) and off-axis model IRDs (bottom) from acoustically measured HRTFs of listeners. Line: normal distribution fitted to the data.

we propose enhancements of the on-axis model aiming at reducing the large IRDs and ANRs in particular when applied to acoustically measured HRTFs of human listeners.

V. DIRECTIONAL OUTLIER REMOVAL

In this section, we address the large ANRs which resulted when the on-axis model was applied to acoustically measured HRTFs. The working hypothesis was that a few artifacts caused the large ANRs and that removing the corresponding directions before fitting the model will result in smaller ANRs.

Generally, artifacts can be defined as outliers in the sense of estimations being numerically distant from the rest of the estimations (Grubbs, 1969). Figure 3 shows the estimated TOAs for an exemplary listener and from that visual inspection, some of the data points might be manually classified as outliers.

There are many methods for detection and removal of outliers from data (Barnett and Lewis, 1994). Most unsupervised methods rely on the assumption of normally distributed data and test individual data points for the probability of deviating from the normality (NIST/SEMATECH, 2002). Grubbs' test is such a test, it is, however, only recommended when testing for a single outlier. When the number of outliers is not exactly known, the generalization of Grubbs' test known as the extreme studentized deviate (ESD) test is recommended (Rosner, 1983). Since we can not specify the exact number of the outliers, the ESD test seems to be the appropriate test for our outlier detection.

The ESD test requires many data points in the pool, thus, TOAs from all available directions of an HRTF set are pooled. Unfortunately, TOAs are correlated with Φ , and thus, if collapsed over Φ and Θ , the TOAs were not normally distributed. This situation is illustrated in Fig. 8, in which the top panels show the TOAs as a function of lateral angle Φ , i.e., collapsed over the polar angles Θ , for HRTFs of STP (left column) and of NH89 from DB1 (right column). The TOAs were obtained from the MCM estimator and black points represent outliers found later by the ESD test (see later). While in that representation some data can be visually identified as outliers, most of them would become undetectable when further collapsing over Φ . Basically, by further collapsing over Φ , the variance of the pool would be large and the outliers would remain hidden among the majority of the other data.

In order to better fulfill the assumption of normality and to obtain a data pool with a smaller variance, the *residuals* resulting from the on-axis model fit are considered. The bottom panels of Fig. 8 show the residuals corresponding to the TOAs shown in the top panels. Note that in that representation, the outliers will remain far from the majority of the data points even when further collapsing over Φ . Thus, the residuals show a better approximation of a normal distribution with a smaller variance, which makes them a more suitable basis for the unsupervised outlier detection with the ESD test.

In the ESD test, the hypothesis of no outliers in the data set is tested at a given significance level up to an

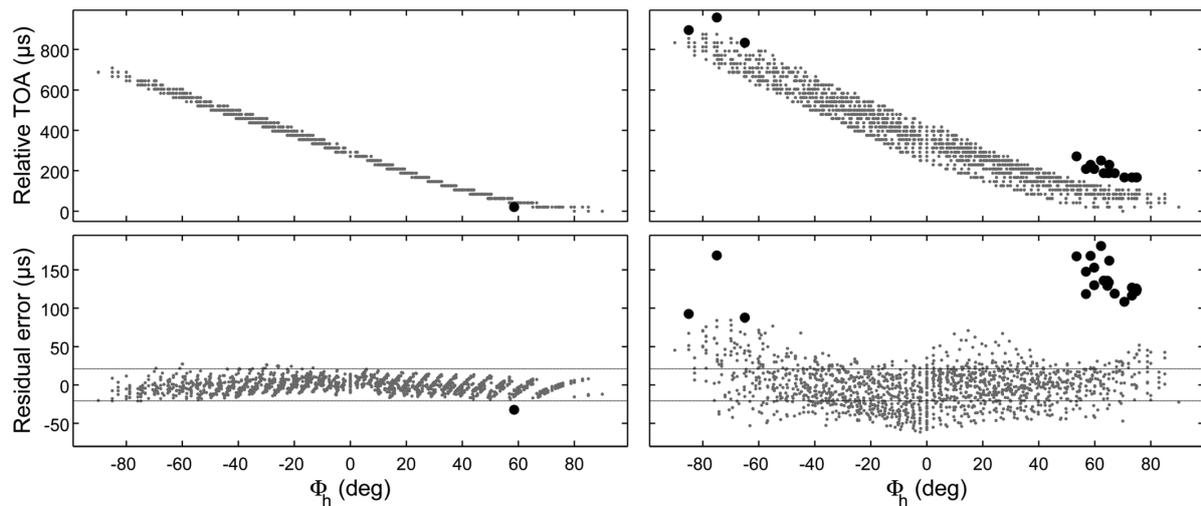


FIG. 8. Relative TOAs (top row) and on-axis model fit residuals (bottom row) from HRTFs of STP (left column) and of an exemplary listener (NH89, DB1; right column). Black points: data classified as outliers by the ESD test with an upper bound of outlier rate of 1% (see text). The reference for the relative TOAs is the smallest TOA in each HRTF set. Horizontal lines: ± 1 sampling interval.

upper bound of outliers (compare Sec. 1.3.5.17.3 in NIST/SEMATECH, 2002). The significance level is usually chosen as 0.05 and was also used in our study. The upper bound can be chosen as a compromise of finding only a few outliers when tested for clean data and but still removing most of the outliers when tested for noisy data. In our study, TOAs obtained for calculated HRTFs of STP represented clean data and TOAs obtained for acoustically measured HRTFs represented noisy data. This comparison is illustrated in Fig. 8. The black points represent outliers resulting from the ESD tested with an upper bound of outlier rate of 1%, the gray points represent the remaining data. From this visual inspection, it seems like the ESD test was able to detect outliers that also in a supervised procedure would have been removed.

In order to estimate an appropriate upper bound of outliers, the outlier rate and ANRs were calculated for the objects Sphere (all conditions from Sec. IV B 1), STP and SAT (from Sec. IV B 2), and human listeners (from Sec. IV C). The MCM estimator and the on-axis model were used as described in Sec. IV C. The on-axis model was fit to TOAs and ANRs were calculated. Then, ESD tests were performed with upper bounds of the outlier rate of 0.01, 0.1, 1, and 10%. The outliers were removed and the on-axis models were fit again, this time based on the outlier-adjusted set of TOAs. Outlier rates and the final ANRs were calculated.

The statistics of the outlier rates and ANRs are summarized in Table IV. For the object Sphere, no outliers were detected in all conditions, and thus, these conditions are not shown. ANRs and outlier rates are positive values and can be approximated by the normal distribution only when their distribution has a small variance. Thus, in addition to the average and standard deviation, the median, and maximum values are shown.

The outlier rate increased with increasing upper bound. For calculated HRTFs (STP and SAT), the outlier removal did not affect the ANRs at all, and it did not affect the outlier

rates up to the upper bound of 1%. For measured HRTFs, the outlier removal had an effect on the ANR and outlier rate even at the smallest upper bound tested, but an increase of the upper bound from 1% to 10% did not substantially change the ANR.

TABLE IV. Outlier rates obtained from the outlier detector and ANRs obtained from fitting the on-axis model to outlier-adjusted TOA sets. UB: upper bound of outlier rates used in the outlier detector (hyphen: without outlier detector). Avg \pm SD: Average \pm 1 standard deviation. Q₂: Median (second quartile). All: Results for measured HRTFs of all listeners.

	UB (%)	ANR (μ s)			Outlier rate (%)	
		Avg \pm SD	Q ₂	Max	Q ₂	Max
SAT	-	6.7	-	-	-	-
	0.01	6.7	-	-	0	-
	0.1	6.6	-	-	0.29	-
	1	6.5	-	-	0.38	-
	10	6.4	-	-	0.81	-
STP	-	8.9	-	-	-	-
	0.01	8.9	-	-	0	-
	0.1	8.9	-	-	0.06	-
	1	8.8	-	-	0.10	-
	10	8.6	-	-	0.68	-
All	-	40.4 \pm 46.5	24.0	407	-	-
	0.01	30.9 \pm 13.6	22.5	89.5	0.00	2.67
	0.1	30.5 \pm 13.5	22.2	89.5	0.06	4.28
	1	29.8 \pm 13.5	21.6	89.5	0.16	6.42
	10	28.2 \pm 13.4	20.4	89.5	0.84	10.7
DB1	-	27.6 \pm 8.7	22.3	53.9	-	-
	1	25.0 \pm 8.9	20.3	53.0	0.32	4.65
DB2	-	56.2 \pm 43.9	38.9	322	-	-
	1	42.8 \pm 18.2	29.0	89.5	0.08	6.42
DB3	-	51.8 \pm 79.1	25.5	407	-	-
	1	25.8 \pm 3.7	23.3	40.8	0.00	6.42

Thus, the upper bound of outlier rate of 1% seems to be a good compromise. It did not have much effect on the calculated HRTFs (clean TOAs) but a significant effect on the measured HRTFs (noisy TOAs). For the latter, the outlier removal substantially improved poor fits (from largest ANR of $407 \mu\text{s}$ to largest ANR of $89.5 \mu\text{s}$), resulting in reduced variability across the listeners (from standard deviation of 46.5 to that of $13.5 \mu\text{s}$). The outlier rate was at a maximum of 6.4% , with a median over all listeners of 0.16% , i.e., statistically below the upper bound of 1% .

VI. OFF-AXIS TOA MODEL

The on-axis model, evaluated for calculated HRTFs did not raise concerns even for HRTFs of a complex object like the STP. However, when applied to acoustically measured HRTFs of human listeners, the estimated IRDs were larger than expected. In this section, we extend the on-axis model in order to address those large IRDs. We analyze the problem, build a hypothesis for the reason of the large IRDs, and test it. Then, a model referred to as off-axis model is proposed. It is supposed to result in smaller IRDs even when applied to acoustically measured HRTFs of human listeners. The model is evaluated and finally applied to acoustically measured HRTFs.

A. Problem analysis

Figure 7 (top panel) shows the histograms of the IRD resulting from the on-axis model fit to HRTFs of human listeners. This histogram represent the statistics of IRDs provided in Sec. IV C. Note particularly the large variability indicating that for some listeners, the IRD was in the range of a few centimeters. For a comparison, the IRDs obtained for the calculated HRTFs of the Sphere (all conditions from Sec. IV B 1), SAT, and STP were much smaller, with an average of 0.0 ± 0.1 mm.

Seeking for an explanation of this observation, we compared the TOAs estimated by the MCM estimator and the TOAs from the on-axis model of an exemplary listener (NH89, DB1). The corresponding estimated model parameters are provided in Table III. Figure 9 shows the corresponding TOAs as functions of the azimuth angle in the interaural horizontal plane. For the left ear, the difference between the smallest and the largest TOA was $784 \mu\text{s}$. This difference would correspond to a maximum ITD of $784 \mu\text{s}$, if the HRTFs measured for the left ear were mirrored to the right side along the median plane. Note that this procedure is sometimes used in order to save time during the measurements (e.g., Gardner and Martin, 1995). For the right ear, the difference between the smallest and the largest TOA was $532 \mu\text{s}$, and thus, the maximum ITD would be $532 \mu\text{s}$. In the light of this finding, it is not surprising that our model converges at interaurally different sphere radii (107.3 mm and 68.5 mm for the left and the right ear, respectively) yielding an interaural asymmetry. Interestingly, the fits shown in Fig. 9 do not raise evidence for a fail of the model for any of the two ears. The asymmetry appears to be an intrinsic property of the measured TOAs and not an artifact of the modeling itself.

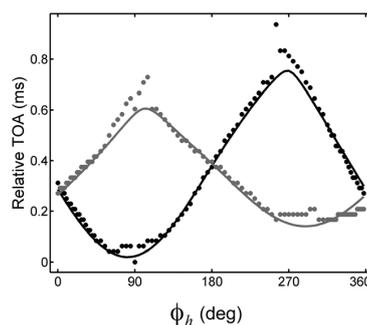


FIG. 9. Relative TOAs of an exemplary listener (NH89, DB1) in the interaural horizontal plane for the left (black) and right (gray) ear as results from the MCM estimator (symbols) and the on-axis model (lines). The reference for the relative TOAs is the smallest TOA in HRTF sets of both ears.

Such an asymmetry might result from a laterally non-centered position of the listener's head in the HRTF measurement. Interestingly, the descriptions of the three HRTF databases claim that the listener was placed at the center of the measurement, i.e., on-axis. In order to investigate how a lateral offset of an object affects the asymmetry of the TOA, the on-axis model was applied to HRTFs calculated for the objects Sphere with radius of 77.5 , 87.5 , and 97.5 mm, each of them placed 0 , 10 , and 20 mm to the left of the center. The other parameters were $\phi_e = \pm 90^\circ$, and $\theta_e = 0^\circ$. The MCM estimator was used to obtain the TOAs. For the fits, the initial values were $r = 90$ mm, $\phi_e = \pm 80^\circ$, and $\theta_e = 10^\circ$. Figure 10 shows the actual and estimated parameters. While the resulting ANR was below $7.3 \mu\text{s}$ in all conditions, the IRDs increased with increasing off-axis position of the object.

This finding suggests that the large IRDs obtained from the on-axis model when applied to acoustically measured HRTFs originate from a potential listener's lateral offset in the measurement. The on-axis model does not consider such an offset and thus results in a large IRD. Thus, in the following, we derive an off-axis model, which allows arbitrary positions of the listener's head. Note that a procedure jointly fitting the on-axis model to TOAs for both ears and forcing the IRD to zero would be an option as well. In such a procedure, however, the ANR would increase as a compromise to fit an average radius to TOAs for both ears. A model allowing arbitrary positions might have the potential to preserve the ANR and still reduce the IRD.

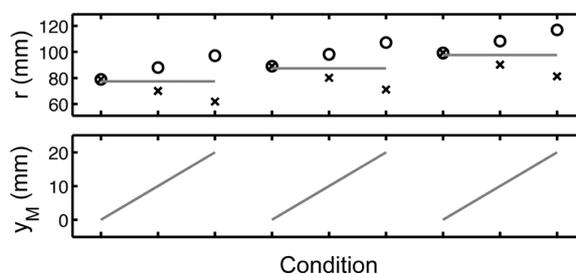


FIG. 10. Estimated on-axis model parameters from HRTFs calculated for the non-centered object Sphere. Other details as in Fig. 5.

B. Model derivation

We consider the spherical head model to be arbitrarily positioned in the measurement system (Fig. 11). In order to consider the arbitrary position, $\vec{M} = [x_M y_M z_M]$ represents the offset of the sphere center relative to the measurement center. Note that while the head model might be arbitrarily positioned, its position still must be stationary, i.e., must not vary over the time. The direction-independent part of the sound propagation path ends on the surface of a centered sphere with radius $r + |\vec{M}|$ (see Fig. 11, bottom panel). The direction-dependent part of the sound propagation path is represented by the sum of three paths s_1 , s_2 , and s_3 , with s_1 from Eqs. (3) and (4). In order to derive s_2 , the angular distance β between the HRTF direction \vec{h} and the position vector \vec{M} is calculated as follows:

$$\beta = \arccos [-\cos \theta_h (x_M \cos \phi_h + y_M \sin \phi_h) - z_M \sin \phi_h]. \quad (6)$$

Then, the sound-propagation path s_2 is given by

$$s_2 = -r + |\vec{M}| \cos \beta + \sqrt{r^2 + 2|\vec{M}|^2 \cos^2 \beta + r|\vec{M}|}. \quad (7)$$

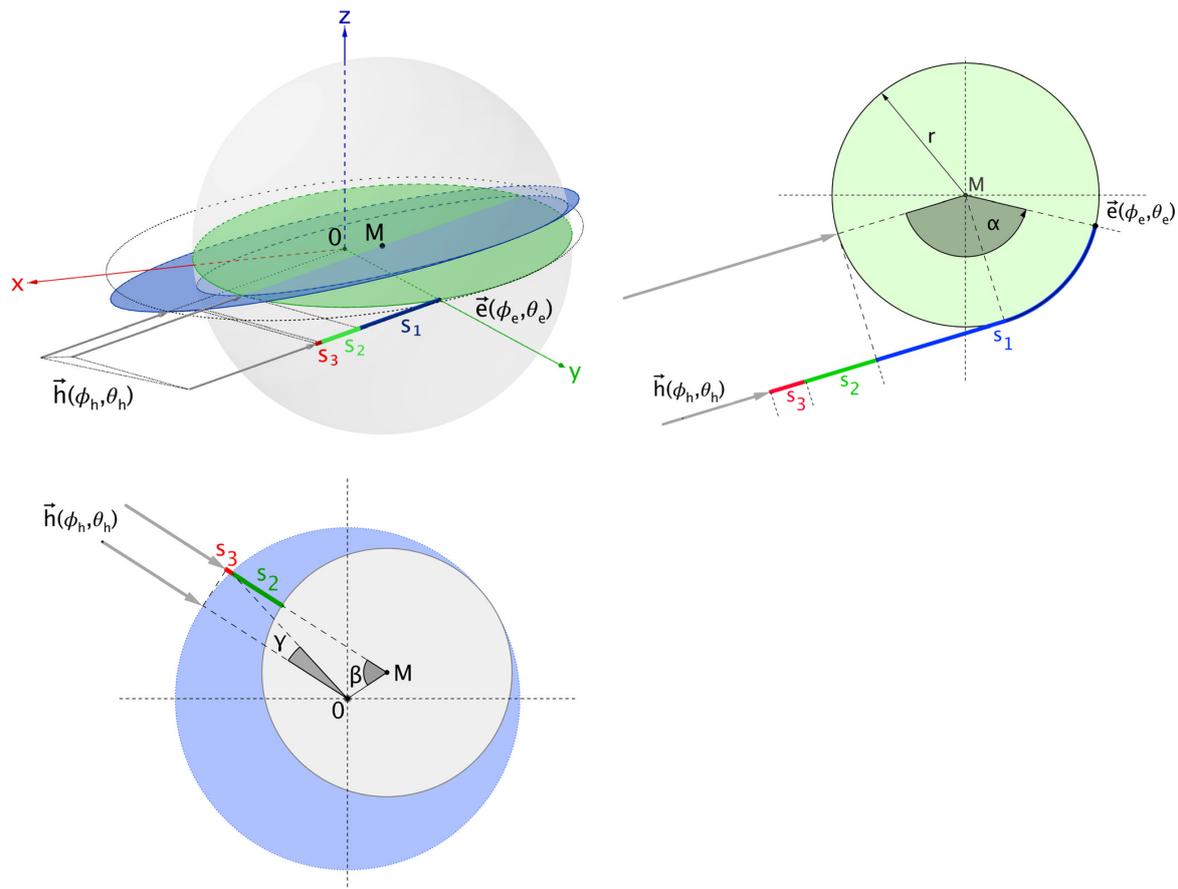


FIG. 11. (Color online) Off-axis model. Top-left panel: Sphere representing the head placed off-axis of the coordinate system. The large circle is the sphere's tangent circumscribed circle with the center in the origin of the coordinate system. Bottom-left panel: Projected view on that tangent circle showing the off-axis sphere and the first two components of the propagation paths. Right panel: Projected view on the great circle of the sphere showing all components of the propagation path. Note that the two projections are not necessarily in the same plane.

Further, the sound-propagation path s_3 is

$$s_3 = \frac{|\vec{M}| \sin \beta}{2(|\vec{M}| + r) \tan \frac{\gamma}{2}}, \quad (8)$$

with

$$\gamma = \pi - \beta - \arccos \left(1 - \frac{2r s_2 + s_2^2}{2|\vec{M}|^2 + 2r|\vec{M}|} \right). \quad (9)$$

Finally, the modeled TOA $\tilde{\tau}$ is

$$\tilde{\tau} = \frac{(s_1 + s_2 + s_3)}{c} + \bar{\tau}. \quad (10)$$

In order to obtain the model parameters \vec{M} , $\bar{\tau}$, r , ϕ_e , and θ_e , Eq. (10) is fit to the set of estimated TOAs $\hat{\tau}$ using the same algorithm as that used for the on-axis model.

The off-axis model has many degrees of freedom and there are many local minima at which the fit might converge. In order to account for the increase in degrees of freedom, the termination tolerance was reduced to 10^{-8} . The initial values, however, require a special handling. If the initial values used for the on-axis model ($r = 87.5$ mm

and $\vec{M} = 0$) were directly used for the off-axis model, the fit might converge at a local minimum at $\vec{M} = 0$ and a radius corresponding to that obtained from the on-axis model, which, again then, would yield an IRD comparable to that obtained from the on-axis model. In order to help the algorithm to converge at a global minimum, we used the estimated parameters from the on-axis model as the initial values for the off-axis model. This was done for ϕ_e , θ_e , and $\bar{\tau}$, but not for \vec{M} (in the on-axis model, \vec{M} was fixed to zero per definition) and not for r (resulted from the on-axis model in a large IRD). In order to still provide reasonable initial values for r and \vec{M} , the initial value for r was the binaural average of the radii obtained from the on-axis model for both ears. The initial value for \vec{M} was set to $\vec{M} = [1 \ 0.5 \cdot \text{IRD} \ 1]$ mm, because \vec{M} resulting from the off-axis model was supposed to show a lateral offset toward the direction of the ear with the larger radius estimated by the on-axis model.

Even with those initial values, the fitting algorithm might still converge at $\vec{M} = 0$ and at r of the on-axis model. In order to prevent that, constraints were used with r at ± 0.25 IRD, $|\vec{M}|$ of ± 10 cm, ϕ_e of $\pm 45^\circ$, θ_e of $\pm 45^\circ$, and $\bar{\tau}$ of ± 1 ms. Note that with that strategy, the initial values for the off-axis model depend on the results from the on-axis model, for which the initial values must be provided.

C. Evaluation

1. General ability

The general ability of the off-axis model to describe the estimated TOAs was evaluated. To this end, the evaluation from Sec. IV B 2 was repeated for the off-axis model. The MCM estimator was applied on HRTFs of the objects Sphere, SAT, and STP. Then, for each object, the on-axis model was fit (initial values: $r = 90$ mm, $\phi_e = \pm 80^\circ$, and $\theta_e = 10^\circ$) and subsequently, the off-axis model was fit (initial values based on the results from the on-axis model). The full sets of TOAs were used, i.e., outliers were not removed.

Table V shows ANRs and estimated parameters. The results are similar to those from the on-axis model, demonstrating the general ability of the off-axis model to describe the estimated TOAs.

TABLE V. Parameters and ANRs (average ± 1 standard deviation) resulting from fitting the off-axis model to TOAs estimated from HRTFs of SAT, STP, and all listeners (All). Full: Fits to full TOA sets, i.e., without outlier removal. O-A: Fits to the outlier-adjusted TOA sets. L: Left ear. R: Right ear.

	TOA set	Ear	r (mm)	ϕ_e ($^\circ$)	θ_e ($^\circ$)	x_M (mm)	y_M (mm)	z_M (mm)	ANR (μ s)
SAT	Full	L	90.3	85.0	-8.5	0.0	-0.2	-2.2	6.5
		R	90.3	-84.9	-8.5	-0.1	0.2	-2.1	6.8
	O-A	L	90.3	84.9	-8.8	0.0	-0.2	-1.8	6.3
		R	90.3	-84.9	-8.8	-0.1	0.2	-1.7	6.6
STP	Full	L	93.2	82.9	-9.3	0.3	0.1	0.6	9.1
		R	93.2	-83.1	-9.3	0.8	0.0	0.9	9.4
	O-A	L	93.2	82.8	-9.3	0.2	0.1	0.6	9.0
		R	93.2	-83.2	-9.4	1.0	0.0	1.0	9.4
All	Full	L	89.6 ± 5.5	85.2 ± 8.4	-9.7 ± 10.7	-4.6 ± 12.2	7.1 ± 13.9	7.1 ± 17.0	39.2 ± 46.3
		R	89.5 ± 5.9	-83.3 ± 9.9	-12.7 ± 7.9	-6.5 ± 12.7	6.6 ± 17.9	9.9 ± 14.0	36.3 ± 40.9
	O-A	L	88.6 ± 5.3	85.2 ± 9.2	-12.3 ± 6.9	-4.3 ± 12.6	7.0 ± 13.3	10.5 ± 12.1	28.4 ± 13.0
		R	88.5 ± 5.5	-84.6 ± 9.3	-13.9 ± 7.4	-4.7 ± 12.3	6.2 ± 17.2	11.0 ± 12.0	28.9 ± 12.5

2. Effect of parameter variation

The off-axis model was evaluated for TOAs estimated from HRTFs obtained under different conditions, represented by spheres of different sizes and different microphone positions. This evaluation corresponds to that from Sec. IV B 1 for the MCM estimator. Correspondingly, the evaluation was performed under similar conditions, with the modification that after fitting the on-axis model, the off-axis model was fit to the full sets of TOAs.

The resulting errors, i.e., differences between the actual and estimated parameters are shown in rows labeled as Centered in the Table VI. The off-axis model described the different conditions with the radius errors of approximately 1.6 mm, the off-axis position errors smaller than 0.4 mm, and the ear position errors of approximately 1.2° . The ANRs ($6.2 \pm 0.1 \mu$ s) were similar to those for the on-axis model ($6.5 \pm 0.1 \mu$ s).

3. Susceptibility to the lateral offset

The susceptibility of the off-axis model to HRTFs of non-centered object is an important evaluation. The on-axis model resulted in large IRDs when applied to such HRTFs and the off-axis model was supposed to result in much smaller IRDs. To this end, the off-axis model was applied to the HRTFs calculated for the object Sphere placed at a combination of following offsets: x_M of 0, -10, and -20 mm; y_M of 0, 10, and 20 mm; and z_M of 0 and -10 mm (compare Sec. VIA). The other parameters were $r = 77.5$, 87.5, and 97.5 mm, $\phi_e = \pm 90^\circ$, and $\theta_e = 0^\circ$. The MCM estimator was used to obtain the TOAs and initial values from Sec. IVD were used. The off-axis models were fit to the full sets of TOAs.

The estimated parameters are shown in Fig. 12. The parameter errors are shown in rows labeled as Non-centered in the Table VI. The IRD was 0.0 ± 0.1 mm on average over all conditions tested.

4. Effect of outlier removal

In this evaluation, the effect of the outlier removal on the off-axis model was evaluated on the calculated HRTFs of SAT, STP, and centered and non-centered object Sphere.

TABLE VI. ANRs and parameter errors (average ± 1 standard deviation) resulting from fitting the off-axis model to TOAs estimated from HRTFs of the object Sphere. Centered: Conditions in which r and \vec{e} varied for $\vec{M} = 0$. Non-centered: Conditions in which \vec{M} varied. Other details as in Table V.

Condition	TOA set	Ear	r error (mm)	ϕ_e error ($^\circ$)	θ_e error ($^\circ$)	x_M error (mm)	y_M error (mm)	z_M error (mm)	ANR (μs)
Centered	Full	L	1.6 ± 0.2	-1.3 ± 0.3	0.9 ± 0.5	-0.2 ± 0.5	0.1 ± 0.2	-0.4 ± 0.9	6.1 ± 0.1
		R	1.6 ± 0.2	1.2 ± 0.4	0.8 ± 0.6	-0.1 ± 0.5	-0.1 ± 0.2	-0.3 ± 0.9	6.1 ± 0.1
	O-A	L	1.6 ± 0.2	-1.3 ± 0.3	0.9 ± 0.5	-0.2 ± 0.5	0.1 ± 0.2	-0.4 ± 0.9	6.1 ± 0.1
		R	1.6 ± 0.2	1.2 ± 0.4	0.8 ± 0.6	-0.1 ± 0.5	-0.1 ± 0.2	-0.3 ± 0.9	6.1 ± 0.1
Non-centered	Full	L	2.2 ± 0.6	-1.3 ± 0.3	0.7 ± 0.3	-0.1 ± 0.4	-0.8 ± 0.6	-0.0 ± 0.5	6.2 ± 0.1
		R	2.6 ± 0.8	1.3 ± 0.3	0.6 ± 0.3	-0.0 ± 0.5	0.8 ± 0.8	0.2 ± 0.6	6.2 ± 0.1
	O-A	L	2.3 ± 0.5	-1.2 ± 0.3	0.7 ± 0.3	0.0 ± 0.4	-0.9 ± 0.5	0.0 ± 0.5	6.2 ± 0.1
		R	2.3 ± 0.8	1.1 ± 0.3	0.5 ± 0.4	0.2 ± 0.5	0.6 ± 0.8	0.3 ± 0.7	6.2 ± 0.1

While these conditions represent clean data, in this evaluation, the potential negative impact of the outlier removal on the off-axis model is investigated. The evaluation corresponds to those from the previous three sections with a single modification: With the results from the on-axis model, the outlier detection was performed with the upper bound of outlier rate of 1% and then the off-axis models were fit to the outlier-adjusted sets of TOAs.

For SAT and STP, the outlier rate was 0.13% and 0.06%, respectively, and the estimated parameters and ANRs are shown in Table V (rows labeled as O-A). An effect of the outlier removal cannot be observed.

For the object Sphere, the parameter errors and ANRs are shown in Table VI. For the centered objects (variation of r and \vec{e} ; rows labeled as Centered), the outlier rate was 0%, and thus, the effect of the outlier removal is not given. For the non-centered objects (variation of \vec{M} , rows labeled as

Non-centered), the average outlier rate was 0.7%. Compared to the full TOA sets, the estimated parameter error change was in the range of 1 mm and 1° , and the ANRs decreased by $0.1 \mu\text{s}$.

In summary, for calculated HRTFs, the outlier removal did not raise any evidence for a significant negative effect on the results of the off-axis model.

D. Application

The off-axis model was applied to HRTF sets of human listeners. This application corresponds to that from Sec. IV C, with the modification of using the off-axis model. The MCM estimator was used to obtain TOAs from acoustically measured HRTFs of human listeners described in Sec. II B 2. Then, for each listener, the on-axis model was fit with the initial values of $r = 87.5 \text{ mm}$, $\phi_e = \pm 90^\circ$, and $\theta_e = 0^\circ$. Further, off-axis models were fit to both full and outlier-adjusted TOA sets.

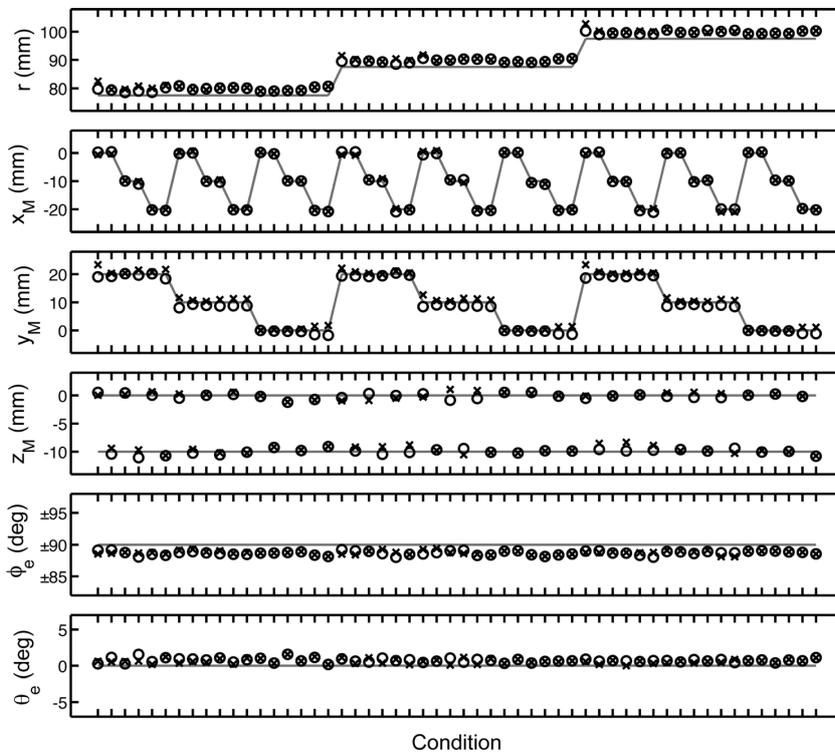


FIG. 12. Estimated off-axis model parameters from HRTFs calculated for the non-centered object Sphere. Other details as in Fig. 5. ϕ_e : Positive and negative values correspond to the left and the right ears, respectively.

The rows labeled as All in Table V show the estimated parameters for the full (rows labeled as Full) and outlier-adjusted (rows labeled as O-A) TOA sets. The outlier removal reduced the variability in the goodness-of-fit, as indicated by the decrease of the ANR standard deviation from $46.3 \mu\text{s}$ (Full) to $12.5 \mu\text{s}$ (O-A). Also the estimated parameters show a slightly reduced variability when the outlier removal was applied.

Figure 7(bottom panel) shows the histogram of the IRDs obtained with the off-axis model fit to the outlier-adjusted TOA sets. The average IRD was $0.05 \pm 1.54 \text{ mm}$ and $-0.14 \pm 0.97 \text{ mm}$, with individual IRDs ranging from -7.1 to 9.9 mm and from -4.2 to 3.8 mm for the full and outlier-adjusted TOA sets, respectively.

E. Discussion

The off-axis model, applied to HRTF calculated for centered objects like SAT, STP, and Sphere, reconstructed the actual geometry parameters with an error up to a small fraction of the sampling interval (of $20.8 \mu\text{s}$ or 7.1 mm). Also for non-centered objects, the model seems to be able to yield correct parameters, as shown in Fig. 12 for various HRTF sets of the non-centered object Sphere. With an average IRD of $0.0 \pm 0.1 \text{ mm}$, the model did not fail for the non-centered objects. Recall that this issue was the main problem of the on-axis model. The average ANR was in the range of a few microseconds, indicating a good fit of the model. The outlier removal did not show any negative effects on the results for calculated HRTFs.

The off-axis model and outlier removal were applied to acoustically measured HRTFs of human listeners. The average and largest outlier rates were 0.7% and 6.4%, respectively. Note that at that point, the estimated parameters can not be compared to the actual parameters of the measured objects, but, they can be checked for plausibility. For example, the estimated radius was $89.2 \pm 5.4 \text{ mm}$ when averaged across both ears and all listeners from the three HRTF databases. Assuming an adequate human sub-population in those databases, the estimated average radius is close to the average head radius of adult humans of 87.5 mm (Hartley and Fry, 1921). Also the estimated average ear positions, nearly most lateral (ϕ_e of $\pm 85^\circ$) and a little below the head center (θ_e of -12°), correspond to those usually found in human listeners (Algazi et al., 2001b; Burkhard and Sachs, 1975).

The main motivation for the off-axis model were the large IRDs obtained with the on-axis model. For acoustically measured HRTFs, the IRDs obtained with the off-axis model ($0.03 \pm 1.12 \text{ mm}$) were much smaller and showed smaller variance than that obtained with the on-axis model ($12.3 \pm 22.9 \text{ mm}$). Thus, the off-axis model seems to fulfill the expectations of decreasing the estimated IRD without increasing the ANR. Together with the outlier removal, the off-axis model was even able to further decrease the ANR (binaural average of $28.6 \pm 12.7 \mu\text{s}$).

Two further issues were observed. First, the application of the on-axis model resulted in a positive average IRD (12.5 mm), which implies a consistent lateral offset in the measurement position. Indeed, the application of the off-axis model resulted in a positive average lateral offset y_M (on

average 6.6 mm), indicating that the listeners were, on average, placed to the left of the center in the measurements in all three HRTF databases.

Second, the radius was consistently overestimated, in both on-axis and off-axis models. On average, the approximate bias was 2.5 mm for the SAT, and 5.7 mm for the STP conditions. The origin of the consistent overestimation of the radius is unclear. One speculative explanation might be the pinna, which might increase the effective propagation delay captured in the HRTFs, but is missing in the TOA model. Actually, an anthropometric model of the ITD showed a significant statistical importance of the pinna protruding height as a parameter of the model (Xie, 2013), showing evidence for the relevance of the longer propagation path with the pinna. In our model, the underestimation of the propagation paths yields an overestimation of the head radius, an issue which might be considered in further improvements of the proposed model.

VII. CONCLUSION

A direction-continuous model for the monaural broadband TOA describing the HRTF measurement geometry was proposed. The model is fit to direction-discrete TOAs resulting from a TOA estimator applied on a set of HRTFs. The first model version, the on-axis model, assumes an exact placement of the listener in the center of the HRTF measurement setup. The on-axis model has been evaluated with calculated HRTFs of objects involving a sphere, torso, and pinna. While it was able to accurately reconstruct the geometric parameters used in calculated HRTFs, it failed when applied to acoustically measured HRTFs of human listeners in terms of large IRDs and ANRs. In the evaluation, four TOA estimators were used. The MCM estimator appeared to be the most-compatible estimator for the proposed model.

Basing on the observation that the large IRDs are linked with an off-axis placement of the object in HRTF simulations, a model able to represent measurement geometry with an arbitrary placement of the listener was proposed. This off-axis model was evaluated with calculated HRTFs of on-axis and off-axis placed objects, resulting in parameter errors smaller than the spatial resolution of the propagation paths. When applied to acoustically measured HRTFs of human listeners, the off-axis model resulted in parameters consistent with realistic listener geometry, particularly, yielding IRDs close to zero.

In order to improve the goodness of fit, an unsupervised outlier removal based on the ESD test was proposed. Together with the off-axis model, the outlier removal resulted in substantially reduced ANRs, particularly improving the problematic fits without any interaction of the operator.

The results from the application of the off-axis model show the relevance of an arbitrary position when modeling TOAs obtained from realistic HRTF measurements in human listeners. Interestingly, distinct lateral offsets were found in HRTFs of most of the listeners in all three HRTF databases tested, indicating that in those measurements lateral shifts have been involved.

The parameters of the off-axis TOA model describe the geometric setup used for the HRTF measurement. The model output is the TOA as a continuous directional function.

The implementations of the TOA models with outlier removal are available in the Auditory Modeling Toolbox (Søndergaard and Majdak, 2013). These implementations allow to further analyze and modify HRTFs. For example, TOAs can be removed from each HRTF, and only the model parameters have to be stored. Such an HRTF set can be stored using shorter impulse responses. Also, the continuous representation of the TOA enables a spatially smooth interpolation of the timing between HRTFs. This is especially interesting for the interpolation along a horizontal plane where broadband timing cues are most salient. For a binaural reproduction of virtual sound sources, the original HRTFs can be resynthesized by calculating the TOA function according to the model parameters and shifting the shortened HRIRs to the correct position in time. By considering the TOAs for each ear separately, such a representation is also suitable for binaural reproduction of multiple sound sources.⁹ There are potentially further applications for the proposed off-axis model, e.g., reconstruction of listener's position in a HRTF measurement,¹⁰ reduction of measured positions in HRTF measurements,¹¹ and pre-processing for spatial decomposition techniques with spherical harmonics (Pollow *et al.*, 2012), where a smooth spatial phase is essential.

Finally, it is important to note that the proposed model relies on a TOA estimator. The perceptual relevance of the MCM estimator and the off-axis model is, however, not clear yet. First, the model should be able to represent TOAs from HRTF sets even not originating from human beings. For example, the model might be applied to HRTFs of apes, humanoid robots, or other systems, which fulfill the requirements on the geometry representation. Second, the perceptual relevance of the TOA estimator should be defined, e.g., the hypothesis that an HRTF with the original TOA, replaced by the modeled TOA can not be discriminated from the HRTF with the original TOA. While in listener-specific HRTFs, the original perceptually relevant TOA is unknown, it still somehow must be removed. This can be done, for example, by using the minimum-phase version of the HRTF, which results in a modification of the excess phase, and unfortunately, human listeners are sensitive to phase (Moore and Glasberg, 1989). A preliminary study controlling for the effects of both the TOA model and the minimum-phase approximation did not show any major problems of the TOA model (Ziegelwanger, 2012). A more rigorous perceptual evaluation of the off-axis TOA model when applied on human HRTFs is required.

ACKNOWLEDGMENTS

We thank Christian Kasess for fruitful discussions on statistics, and Bernhard Laback and Wolfgang Kreuzer for their comments on a draft of this paper. This study was supported by the Austrian Science Fund (FWF, P 24124-N13) and the University of Music and Performing Arts, Graz.

NOMENCLATURE

α	Angular distance between a sound source and an ear
β	Angular distance between a sound source and \vec{M}

c	Speed of sound
γ	Auxiliary angle for the off-axis model
\vec{e}	Position of the ear on the sphere, $\vec{e} = [\phi_e, \theta_e]$
ϕ	Azimuth angle (general)
ϕ_e	Azimuth angle of the ear on the sphere
ϕ_h	Azimuth angle of the incoming sound
Φ	Lateral angle (general)
h	Head-related impulse response (HRIR)
h_{min}	Minimum-phase version of h
\vec{h}	Direction of the incoming sound, $\vec{h} = [\phi_h, \theta_h]$
θ	Elevation angle (general)
θ_e	Elevation angle of the ear on the sphere
θ_h	Elevation angle of the incoming sound
Θ	Polar angle
\vec{M}	Center of the sphere, $\vec{M} = [x_M, y_M, z_M]$
r	Radius of the sphere
s_1, s_2, s_3	Sound propagation paths
τ	Time of arrival (TOA)
$\hat{\tau}$	Direction-discrete τ resulting from an estimator
$\tilde{\tau}$	Direction-continuous τ resulting from a model
$\bar{\tau}$	Direction-independent τ
x_M, y_M, z_M	Cartesian coordinates of \vec{M}

¹While the ITD is a frequency-dependent quantity with larger values at low frequencies (Kuhn, 1977), it can be sufficiently approximated by a frequency-independent quantity as indicated for a clothed manikin (Kuhn, 1977) or human listeners (Kulkarni *et al.*, 1999). Thus, in this study, both TOA and ITD are considered being frequency independent.

²Blender 2.64, available from <http://www.blender.org> (date last viewed July 8, 2013).

³Available from <http://www.makehuman.org> (date last viewed July 8, 2013).

⁴Available from <http://www.kfs.oeaw.ac.at/hrtf/database/amt/ziegelwan-ger2013.zip> (date last viewed August 4, 2012).

⁵Retrieved from <http://sofacoustics.org/data/database/ari/> (date last viewed January 12, 2012), originating from the “(short)” HRTF set from <http://www.kfs.oeaw.ac.at/hrtf> (date last viewed January 12, 2012).

⁶Retrieved from <http://sofacoustics.org/data/database/cipic/> (date last viewed January 11, 2012), originating from <http://interface.cipic.ucdavis.edu/sound/hrtf.html> (date last viewed January 11, 2012).

⁷Retrieved from <http://sofacoustics.org/data/database/listen/> (date last viewed January 2, 2012), originating from the “compensated” HRTF set from <http://recherche.ircam.fr/equipements/salles/listen/download.html> (date last viewed January 11, 2012) with the sampling rate set to 48 kHz.

⁸MATLAB R2011a, available from <http://www.mathworks.de/products/matlab/> (date last viewed July 8, 2013).

⁹In some applications of virtual acoustics, in order to increase the rendering efficiency, a binaural pair of HRTFs is modeled as a pair of minimum-phase impulse responses and an ITD. While the ITD is a relative quantity and thus sufficient for the description of an acoustic scene with a single sound source, the inter-source time difference is important for the acoustically correct rendering of multiple sound sources in a scene. The inter-source time difference can only be derived when the absolute delay arising from the acoustic wave propagation path from each source to a listener's ear is available, i.e., the TOA. The inter-source time difference can not be derived from the ITD. The proposed model describes the TOA as a function of direction, showing potential for an efficient rendering of multiple sound sources.

¹⁰If the actual listener position can be reconstructed after the HRTF measurements, the actually measured HRTF directions can be calculated and corrected. Then, the exact placement of the listener in the center of the measurement setup will be not required.

¹¹For a reduction of the measurement positions, a direction-continuous model of HRTFs is required. The proposed model addresses the direction-continuous timing only, and the direction-continuous spectral magnitude is still to be solved.

- Algazi, V. R., Avendano, C., and Duda, R. O. (2001a). "Estimation of a spherical-head model from anthropometry," *J. Audio Eng. Soc.* **49**, 472–479.
- Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001b). "The CIPIC HRTF database," in *Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics* (New York, NY), pp. 99–102.
- Bajorski, P. (2011). *Statistics for Imaging, Optics, and Photonics* (John Wiley and Sons, Hoboken, NJ), pp. 220–227.
- Barnett, V., and Lewis, T. (1994). *Outliers in Statistical Data* (John Wiley and Sons, Hoboken, NJ), pp. 1–616.
- Blauert, J. (1997). *Spatial hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA), pp. 1–494.
- Burkhard, M. D., and Sachs, R. M. (1975). "Anthropometric manikin for acoustic research," *J. Acoust. Soc. Am.* **58**, 214–222.
- Busson, S., Nicol, R., and Katz, B. F. G. (2005). "Subjective investigations of the interaural time difference in the horizontal plane," in *118th Convention of the Audio Engineering Society* (Barcelona, Spain), Convention Paper 6324.
- Coleman, T. F., and Li, Y. (1996). "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM J. Optimiz.* **6**, 418–445.
- Constan, Z. A., and Hartmann, W. M. (2003). "On the detection of dispersion in the head-related transfer function," *J. Acoust. Soc. Am.* **114**, 998–1008.
- Defrance, G., Daudet, L., and Polack, J.-D. (2008). "Finding the onset of a room impulse response: Straightforward?," *J. Acoust. Soc. Am.* **124**, 248–254.
- Dennis, J. E., Jr. (1977). "Nonlinear least-squares," in *State of the Art in Numerical Analysis*, edited by D. Jacobs (Academic Press, Waltham, MA), pp. 269–312.
- Gardner, W. G., and Martin, K. D. (1995). "HRTF measurements of a KEMAR," *J. Acoust. Soc. Am.* **97**, 3907–3908.
- Grubbs, F. E. (1969). "Procedures for detecting outlying observations in samples," *Technometrics* **11**, 1–21.
- Hartley, R. V. L., and Fry, T. C. (1921). "The binaural location of pure tones," *Phys. Rev.* **18**, 431–442.
- Huopaniemi, J., and Smith, J. O. (1999). "Spectral and time-domain preprocessing and the choice of modeling error criteria for binaural digital filters," in *Proceedings of the 16th International Conference on Spatial Sound Reproduction of the Audio Engineering Society* (Rovaniemi, Finland), 301–312.
- Jeppesen, J., and Møller, H. (2005). "Cues for localization in the horizontal plane," in *118th Convention of the Audio Engineering Society* (Barcelona, Spain), Convention Paper 6323.
- Jot, J. M., Larcher, V., and Warusfel, O. (1995). "Digital signal processing issues in the context of binaural and transaural stereophony," in *98th Convention of the Audio Engineering Society* (Paris, France), Convention Paper 3980.
- Knapp, C., and Carter, G. C. (1976). "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust. Speech Sig. Proc.*, **ASSP-24**, 320–327.
- Kreuzer, W., Majdak, P., and Chen, Z. (2009). "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," *J. Acoust. Soc. Am.* **126**, 1280–1290.
- Kuhn, G. F. (1977). "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.* **62**, 157–167.
- Kulkarni, A., Isabelle, S. K., and Colburn, H. S. (1999). "Sensitivity of human subjects to head-related transfer-function phase spectra," *J. Acoust. Soc. Am.* **105**, 2821–2840.
- Lee, B., Choi, J., and Kim, M. (2009). "Interaural time difference estimation using cross-power spectrum phase with zerocrossings," in *4th International Conference on Autonomous Robots and Agents (ICARA)*, 222–226.
- Macpherson, E. A., and Middlebrooks, J. C. (2002). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *J. Acoust. Soc. Am.* **111**, 2219–2236.
- Majdak, P., Goupell, M. J., and Laback, B. (2010). "3-D localization of virtual sound sources: effects of visual environment, pointing method, and training," *Atten. Percept. Psychophys.* **72**, 454–469.
- Mehrgardt, S., and Mellert, V. (1977). "Transformation characteristics of the external human ear," *J. Acoust. Soc. Am.* **61**, 1567–1576.
- Michalopoulou, Z.-H., and Jain, R. (2012). "Particle filtering for arrival time tracking in space and source localization," *J. Acoust. Soc. Am.* **132**, 3041–3052.
- Mills, A. W. (1958). "On the minimum audible angle," *J. Acoust. Soc. Am.* **30**, 237–246.
- Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (1995). "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.* **43**, 300–321.
- Moore, B. C. J., and Glasberg, B. R. (1989). "Difference limens for phase in normal and hearing-impaired subjects," *J. Acoust. Soc. Am.* **86**, 1351–1365.
- Morimoto, M., and Aokata, H. (1984). "Localization cues in the upper hemisphere," *J. Acoust. Soc. Jap.* **5**, 165–173.
- Nam, J., Abel, J., and Smith, J. O. III. (2008). "A method for estimating interaural time difference for binaural synthesis," in *125th Convention of the Audio Engineering Society* (San Francisco, CA), Convention Paper 7612.
- NIST/SEMATECH (2002). "e-Handbook of Statistical Methods," <http://www.itl.nist.gov/div898/handbook/> (date last viewed September 27, 2013)
- Oppenheim, A. V., Schaffer, R. W., and Buck, J. R. (1999). *Discrete-Time Signal Processing* (Prentice Hall, Upper Saddle River, NJ), 904 pages.
- Pollow, M., Nguyen, K.-V., Warusfel, O., Carpentier, T., Müller-Trapet, M., Vorländer, M., and Noisternig, M. (2012). "Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition," *Acta Acust. Acust.* **98**, 72–82.
- Raykar, V. C., Duraiswami, R., and Yegnanarayana, B. (2005). "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *J. Acoust. Soc. Am.* **118**, 364–374.
- Rosner, B. (1983). "Percentage points for a generalized ESD many-outlier procedure," *Technometrics* **25**, 165–172.
- Savioja, L., Huopaniemi, J., Lokki, T., and Väänänen, R. (1999). "Creating interactive virtual acoustic environments," *J. Audio Eng. Soc.* **47**, 675–705.
- Søndergaard, P., and Majdak, P. (2013). "The auditory-modeling toolbox," in *The Technology of Binaural Listening*, edited by J. Blauert (Springer, Berlin, Germany), pp. 19–44.
- Strutt, J. W. (1894). *The Theory of Sound* (Macmillan, London, UK), 510 pages.
- Tribolet, J. (1977). "A new phase unwrapping algorithm," *IEEE Trans. Acoust. Speech Sig. Proc.* **25**, 170–177.
- Wang, L., Yin, F., and Chen, Z. (2009). "A hybrid compression method for head-related transfer functions," *Appl. Acoust.* **70**, 1212–1218.
- Wightman, F. L., and Kistler, D. J. (1989). "Headphone simulation of free-field listening I: Stimulus synthesis," *J. Acoust. Soc. Am.* **85**, 858–867.
- Wightman, F. L., and Kistler, D. J. (2005). "Measurement and validation of human HRTFs for use in hearing research," *Acta Acust. Acust.* **91**, 429–439.
- Xie, B. (2013). *Head-Related Transfer Function and Virtual Auditory Display* (J. Ross Publishing, Plantation, FL), 482 pages.
- Ziegelwanger, H. (2012). *Modell zur effizienten Kodierung von Signallaufzeiten für die binaurale Wiedergabe virtueller Schallquellen* (University of Music and Performing Arts, Graz, Austria), pp. 1–112.

Chapter 4

Calculation of listener-specific HRTFs: Effect of mesh quality

This work was published as:

Ziegelwanger, H., Reichinger, A., and Majdak, P. (2013). "Calculation of listener-specific head-related transfer functions: Effect of mesh quality," in *POMA*, **19**, 050017. [doi:10.1121/1.4799868](https://doi.org/10.1121/1.4799868)

This study presents a result of a collaborative research which was based on the initial idea of the third author. The second author developed the geometry acquisition technique, captured the head and pinnae geometry of the subject, and generated the reference and the BASE mesh. I numerically calculated and acoustically measured the HRTFs, evaluated the HRTFs based on model predictions, and generated figure 3. While the first and second author wrote the initial draft (each about their respective contribution), the third author helped finalizing the manuscript.

Co-author 1:

Co-author 2:

Proceedings of Meetings on Acoustics

Volume 19, 2013

<http://acousticalsociety.org/>



ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Psychological and Physiological Acoustics
Session 1pPPa: Binaural Hearing and Binaural Techniques I

1pPPa9. Calculation of listener-specific head-related transfer functions: Effect of mesh quality

Harald Ziegelwanger*, Andreas Reichinger and Piotr Majdak

*Corresponding author's address: Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, Vienna, 1040, Vienna, Austria, harald.ziegelwanger@oeaw.ac.at

The geometry of head and ears defines the listener-specific directional filtering of the incoming sound. The filtering is represented by the head-related transfer functions (HRTFs), which provide spectral features relevant for the localization of sound-sources. HRTFs can be acoustically measured or numerically calculated based on a geometric representation of the listener. While the acoustically measured HRTFs usually provide localization performance similar to that obtained in free-field listening, the performance obtained with numerically simulated HRTFs, however, heavily depends on the quality of the geometric and acoustic model of the listener used for the simulation. In this study, we show how to calculate listener-specific HRTFs with spectral features similar to that from acoustically measured HRTFs for the entire audible frequency range. We review the boundary-element method coupled with the fast-multipole method and we present details on the prerequisites like the geometry- capture technique, acoustical parameters, and the numerical algorithms. Further, the effect of the mesh quality on the HRTFs was investigated by systematically varying the average edge length from 2-5mm. The HRTF amplitude spectra were analyzed and evaluated by visual comparison and in a localization model. The results of the coarser meshes show indications for the required edge length in HRTF calculations.

Published by the Acoustical Society of America through the American Institute of Physics

INTRODUCTION

Head-related transfer functions (HRTFs) describe the direction-dependent filtering of the incoming sound at the entrance of the ear canal due to scattering, shadowing, and reflection caused by the head, torso, and the pinna (Blauert, 1997). Thus, HRTFs depend on the particular details of a listener's geometry. In the frequency range up to 16 kHz, HRTFs contain listener-specific spectral features, which are required for accurate sound source localization in binaural reproduction systems (Møller, 1992). As an alternative to the common acoustic measurement of the HRTFs (Majdak et al., 2007) where small microphones are placed into the ear canals, the HRTF simulation, where HRTFs are numerically calculated from a representation of a listener's geometry, offers potential for a consumer-suitable, non-contact HRTF acquisition method. A lot of research effort has been done on the calculation of HRTFs by the use of the boundary element method (BEM) in the last years (Katz, 2001a, 2001b; Kahana and Nelson, 2006, 2007; Gumerov and Duraiswami, 2009; Kreuzer et al., 2009). However, HRTFs yielding a localization performance as good as that for acoustically measured ones have not been calculated yet. The results of the HRTF simulation heavily depend on the quality of the listener's geometry data, i.e., three-dimensional mesh, which results from the scanning procedure and the post-processing required for the numerical calculations. Particularly, the mesh accuracy and the mesh resolution, i.e., the number of elements in relation to the frequency of interest, are important properties. Also acoustical parameters like the impedance model of hair and skin have been investigated (Treeby et al., 2007a, 2007b, 2007c). However, requirements on these geometry properties and acoustic parameters for HRTF calculation have not been clarified yet.

In this study we aimed at calculating listener-specific HRTFs similar to that obtained with the acoustical measurement procedure and yielding a localization performance as good as that observed in free-field listening. We assumed, that the accuracy of the mesh is vital and the impedance of the hair can be neglected. Thus, in order to reach this goal it was required to work with the best possible mesh quality and accuracy, regardless of the required effort. While a method suitable for possible future consumers should be as fast, convenient, and pleasant as possible, these restrictions do not hold for this test case. Capturing the human pinna was of special interest. Spectral features, which are indispensable for accurate sound localization, are caused by the individual shape of a listener's pinna, in particular, by its folds, undercuts, and deep cavities. However, these geometric details exacerbate the scanning procedure. Parts of some pinnae are not even visible from the outside, which makes surface scanning methods (e.g., laser scanner) unsuitable. Thus, we propose a geometry capturing procedure, which overcomes these difficulties, evaluate the calculated HRTFs and investigate the effect of the mesh resolution on sound source localization.

METHODS

HRTF Measurement

HRTFs were acoustically measured for 1550 directions in a semi-anechoic chamber with the blocked ear-canal method (Møller et al., 1995). The subject was seated in the center of a circular loudspeaker array (see Fig. 1a). To reduce the intensity of reflections the array construction was covered with acoustic damping material. HRTFs were measured with an exponential frequency sweep (0.05 to 20 kHz). The total time required to measure the HRTFs was decreased by applying the multiple exponential sweep method (Majdak et al., 2007). In-ear-microphones (Sennheiser KE-4-211-2), which were placed in the subject's ear canals, were connected via amplifiers (RDL FP-MP1) to the digital audio converters. The subject was rotated horizontally to measure HRTFs at several elevations at once for one azimuth by playing the sweeps and recording the acoustic signals with the microphones. The horizontal and vertical range was 360° and -30° to 80° , respectively. The horizontal resolution was 2.5° and 5° for directions inside and outside of $\pm 45^\circ$, respectively. The vertical resolution was 5° . The position and orientation of the head was tracked during the measurement to observe and reduce head movements.

Equipment transfer functions were derived from a reference measurement in which the in-ear microphones were placed in the center of the array and impulse responses were measured for all loudspeakers. The equipment transfer functions were cepstrally smoothed and their phase spectrum was set to the minimum phase (Oppenheim et al., 1999). The resulting minimum-phase equipment transfer functions were removed from the HRTFs by spectral division. Then, directional transfer functions (DTFs) were calculated using a method similar to the procedure of Middlebrooks and Green (1990). Finally, the directional impulse responses, i.e., the inverse Fourier transforms of the DTFs, were windowed with an asymmetric Tukey window (fade in of 0.25 ms and fade out of 1 ms) to a 5.33-ms duration (256 Samples at sampling rate of 48 kHz). The resulting DTFs are shown in Fig. 3 (AC).

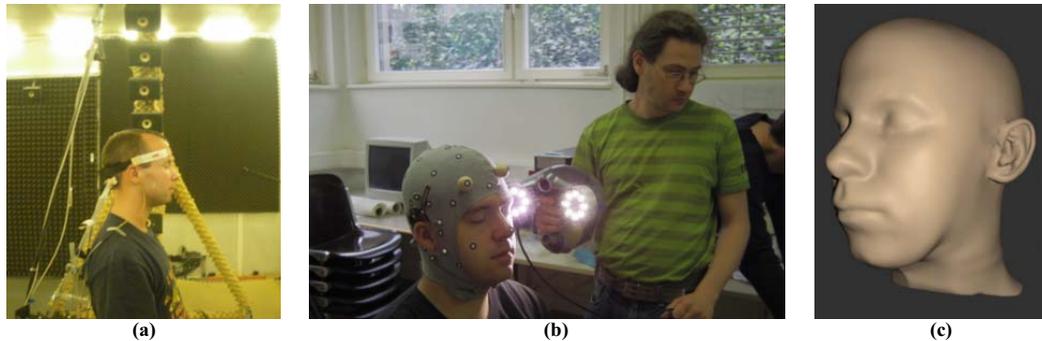


FIGURE 1. (a) Subject during the acoustic HRTF measurements. (b) Subject during the visual scan of the head. (c) Reference mesh of the subjects head with approximately 2.5 million triangular elements.

Geometry Acquisition

Pinna

In order to achieve the most detailed representation of the pinna, a CT scanner was used. Since CT is a volumetric scanning method, the scan is not affected by undercuts and occlusions. The medical scanning-devices typically have too poor resolution and a too high noise level for our purposes. In addition, all CT scanners expose the subject to harmful x-ray radiation. The required scan quality, however, can be achieved with an industrial high-energy CT scanner, where the amount of radiation is not limited by medical concerns. To prevent the subjects from the radiation we made impressions of the ears and scanned them instead of the subjects. Note that other volumetric scanning methods like the magnet-resonance tomography might have been used, however, these methods usually yield global distortions.

The impressions were made of silicone. In order to prevent deformation of the pinna and to capture possibly all of the surface, we used an addition-crosslinked precision silicone with a very low viscosity (viscosity: 2200 MPa·s, shore hardness: A 22). The subjects sealed their ear canals with silicone ear protectors¹ to prevent the silicone from reaching the ear drum. The silicone was pressed-in as deep as possible without being inconvenient for the subject, at least some millimeters past the ear canal opening. The subjects lied on the side, their hair was covered with Vaseline, a cut paper cup was placed over their ears, sealed with modeling clay, and the freshly prepared silicone was slowly filled in with a constant thin flow (see Fig. 2a). The ear was moved to release any large air bubbles captured in cavities. This way, only very small air bubbles remained in the silicone. The silicone material takes about half an hour to harden. After this time, the impression was carefully pulled off the ear, and set aside a day to fully harden. Torn hair was removed from the impression, and it was cleaned from rests of modeling clay. Thin unstable parts were cut resulting in a sufficiently rigid impression.

By using the impressions, the hair on the ears was fully ignored, and only the skin surface was impressed. The hair was either pressed away, or formed very narrow canals that did not affect the CT-scan results. The local surface was very accurate, and some effort was put to also reduce global deformations. The silicone material initially had a very high viscosity leaving the ear in its natural position. After hardening it became hard but elastic (see Fig. 2b).

The impressions were scanned with an industrial CT scanner (v|tome|x c²) using 1000 x-ray images spaced in 0.36° intervals (tube settings: 120 kV, 260 μA, average of three images with 333-ms exposure time each). During the scans, we avoided any stress on the material. The impressions were placed horizontally on the scanner with its flat side on the platform. No other attachments were required. The scanning platform moved abruptly, but the scanner was set to wait 333-ms before the scanning in order to let material oscillations decay.

The resulting volumetric data sets had a cubic voxel size of 50 to 60 μm, totaling to approx. 2 Gigavoxels with 16-bit integer density values. The transition between air and silicone had a width of typically 3 voxels (see Fig. 2c). The isosurface separating air and silicone were extracted using auto-thresholding (VGStudio Max, Volume Graphics GmbH³). This method is a proprietary version of the marching cubes algorithm (Lorenson and Cline, 1987) that extracts the isosurface with sub-voxel precision and curvature-dependent increase in resolution. The effective accuracy can be estimated to be in the range of two to four voxels, i.e., 0.1 to 0.2 mm. From the resulting mesh, we manually removed the surface part produced by the cup wall and the top surface.

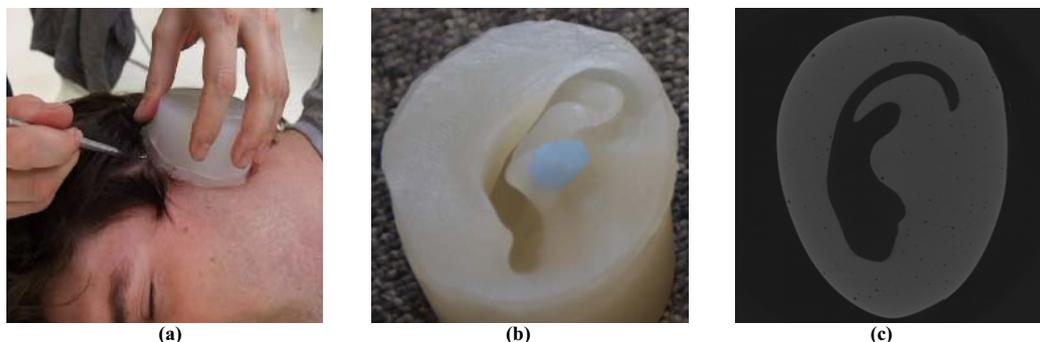


FIGURE 2. (a) Preparation of the silicone impression. (b) Hardened impression. (c) Single layer of CT-scanned impression.

Despite the careful handling of the silicone, air bubbles were inevitable. Most of the bubbles occurred inside the silicone and have been removed by keeping the largest connected surface. However, several bubbles were very close to the surface, leaving a too thin layer of silicone for the CT resolution to find the correct isosurface. In addition, some bubbles actually were on the surface. At these locations, the ear geometry could not be recovered. Most of the small bubbles have been removed based on a curvature threshold that detects the sharp connections between the bubbles and the ear surface (algebraic point set surface, APSS, MeshLab⁴). The remaining bubbles that were not removed with this procedure, have been manually cut and all holes have been filled (Automatic filling method in curvature mode, Geomagic, Geomagic Studio⁵).

Head

The head was scanned by a laser scanner (ZScanner 700CX⁶, ZCorp) projecting two orthogonal laser lines which are manually swept over the scanned surfaces (see Fig. 1b). This scanner features a built-in optical tracking system based on reflective points that are placed directly on the scanned object. Therefore, the scanned parts are always registered relative to the head, ensuring that a globally consistent and accurate model is created, even when the head is moving. We used the highest resolution settings and a tightly set working volume around the head. In order to cope with hair that cannot be scanned, the subjects wore a custom tailored cap with holes for the ears. The mesh, exported from the scanning software, was apparently extracted from an internal volumetric representation with approximately 0.5-mm voxel size and with holes where data could not be captured. These holes were filled with the curvature-based filling algorithms (Geomagic), and the neck was cut and closed.

HRTF Calculation

Free-field HRTFs were calculated using the collocation boundary element method (BEM) with constant elements. To avoid irregular frequencies we used the approach of Burton and Miller (1971). To calculate HRTFs up to 20 kHz in an appropriate time we coupled the BEM with the fast-multipole method (Kreuzer et al., 2009). Calculation time was further decreased by applying the principle of reciprocity (Morse and Ingard, 1987), where the receiver is exchanged with the transmitter. Computation was done on a Linux cluster containing eight machines with Intel i7-3820 processors running at 3.6 GHz and 64 Gbyte of RAM each.

BEM requires a watertight mesh of the head scan and the two pinnae. The pinna meshes were aligned to the pinnae of the head mesh (Manual Registration, Geomagic) using manually placed tie points, and refined (Best Fit Alignment, Geomagic) using only that part of the pinna as alignment reference, which was visible from the side. This was the most reliable part since it was best captured by the surface scanning method of the head, and thus it ensures that the acoustically relevant parts are oriented most accurately. The scan results differed noticeably in the joint region of the head behind the pinna. This was mostly caused by the cap used during the head scan, which slightly enlarged the head. Also a slight rotation of the pinna between both scans might have contributed to the differences. We gave precedence to the pinna scan, since it captured the hard-to-see region behind the ear most detailed and remodeled the head scan to fit seamlessly. To this end, we removed the pinna region of the head scan, and stitched the pinna scans. The seam-region was re-sculpted and smoothed (Voxel-Sculpting, 3D-Coat⁷) in a

resolution of below 0.5 mm voxel-size, taking care to only modify the head mesh. Figure 1c shows the final high accuracy reference mesh, which consisted of approximately 2.5 million triangles. In the following, this mesh is referred to as the base mesh.

Because of memory size, the number of elements used in the BEM was limited to approximately 120 000 elements. Thus, the resolution of the reference mesh was reduced in three steps. First the whole mesh was re-meshed to 1 mm side-length on average ensuring approximately equilateral triangles (Remesh, Geomagic). Then everything except the pinna region was re-meshed to 1.5 mm, the pinna region expanded by 4 polygon rows and the rest further re-meshed to 2.5 mm. The resulting mesh (see Fig. 3a BASE) had approximately 1 mm side-length at the pinna region, 2.5 mm at the head and a transition region with 1.5 mm side-length that was additionally relaxed to reduce degenerate non-equilateral triangles at the region seams. Therefore, most of the polygon budget (approx. 45%) was used for the pinna region, where it is acoustically most relevant, and was reduced continuously to the acoustically less important head region. For acoustic BEM simulations the average edge length of the underlying mesh should lie between a one-tenth and one-sixth of the observed wavelength (Marburg, 2002). Thus, for the numerical HRTF calculation up to 20 kHz an average edge length between 1.7 and 2.8 mm is preferred. Our base mesh fulfills these requirements for the hole boundary surface. To investigate the effect of increasing the average edge length, the surface was uniformly re-meshed to an average edge length of 2, 3, 4, and 5 mm (Openflipper[®], Möbius and Kobbelt, 2012).

The microphone was modeled as sound source by applying a corresponding velocity boundary condition at the blocked entrance of the ear canal. For the rest of the head and pinna the normal velocity was set to zero, i.e., the boundary surface was modeled fully reflective. The impedance of the hair was neglected. The complex sound pressure was calculated at 200 sampling points in the frequency domain, which were linearly spaced between 100 and 20 000 Hz. HRTFs were calculated for 16 022 directions at a distance of 1.2 m. The resolution was 2° for the lateral and the polar angle. After the inverse Fourier transform, we obtained the head-related impulse responses (HRIRs), which were re-sampled at a sampling rate of 48 kHz. Finally, DTFs were calculated following the same procedure as for the acoustically measured HRTFs.

RESULTS

First, the calculated DTFs for the base mesh were evaluated by visual comparison of the resulting amplitude spectra to the acoustically measured data. Figure 3c shows the amplitude spectra in the median plane for the acoustically measured (AC) and calculated (BASE, 2 to 5 mm) DTFs of subject's left ear. Note, that the acoustically measured DTFs were band-limited at 18 kHz and the plot is affected by the coarser spatial sampling in the acoustic measurement setup. The dark parts in the amplitude spectra represent notches, which are the relevant features for sound source localization in the median-plane. They are caused by interference of superposed pinna reflections at the entrance of the ear canal. The fine horizontal lines, which look like torso reflections (we have not included the torso in the mesh) might be caused by the abrupt cutting of the mesh at the bottom end of the neck without any smoothing. Overall, the acoustically measured DTFs (AC in Fig. 3) and DTFs calculated for the base mesh (BASE in Fig. 3) seem to be similar.

A more psychoacoustic-related way to evaluate the calculated DTFs is to analyze the sound localization performance in a localization experiment where a subject listening to sounds filtered with the calculated DTFs responses to the perceived direction. The localization performance can be described by the amount of front-back confusion, i.e., the quadrant error rate, and the local error, i.e., the local polar RMS-error (Middlebrooks, 1999). In our study, the localization performance has been modeled in a sagittal-plane localization model (Baumgartner et al., 2013), which considers listening to sounds filtered with non-individualized DTFs like our calculated ones. The model results are two-fold: 1) response probability of pointing to a direction, and 2) the localization performance parameters, i.e., quadrant error and polar RMS error. Figure 3d shows the response probability to a direction for given polar angles of a virtual sound source, in a simulated localization experiment (baumgartner2013 from the AMToolbox[®]; (Søndergaard and Majdak, 2013)) with the measured and calculated DTFs. The localization model considered a frequency range from 5 to 18 kHz. For the measured DTFs the model predicted a polar error of 29° and a quadrant error rate of 9.3 %. For the base-mesh DTFs, the model predicted increases in the polar and quadrant error of 1° and 1.7 %, respectively. We assume that such small differences can be neglected. Thus, we consider the base-mesh DTFs as perceptually valid.

Further, the HRTFs calculated using meshes with less elements were analyzed. The details on the average edge length, number of elements, and computation time are shown in Table 1. The calculated HRTFs are shown in Fig. 3b and 3c (2 mm to 5 mm). Note the less distinct notches between 5 and 15 kHz for larger average edge lengths of the

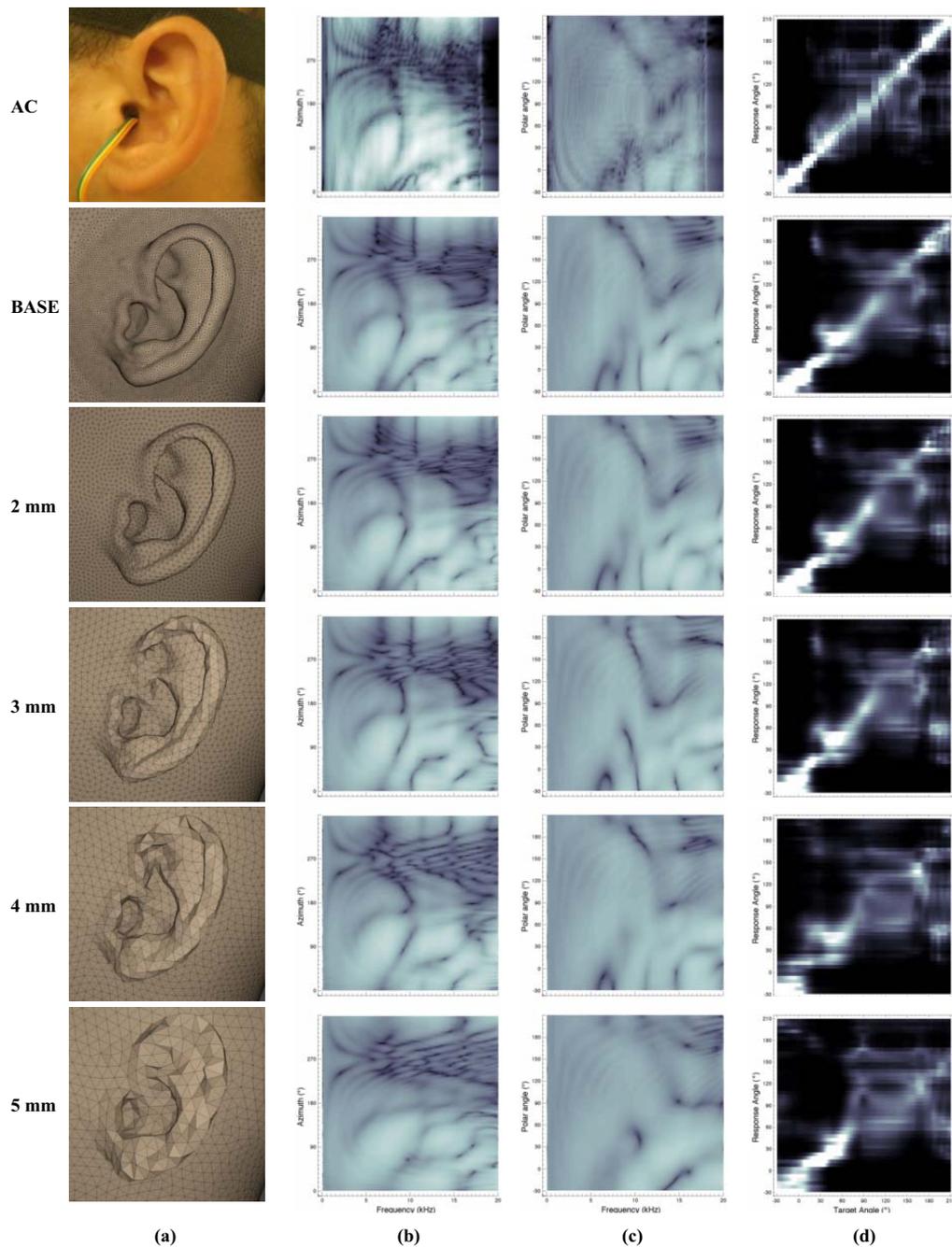


FIGURE 3. (a) Pinna meshes. (b) DTF amplitude spectra for the interaural horizontal plane (brightness: magnitude in dB). (c) DTF amplitude spectra for the median plane. (d) Response probability matrix (brightness: response probability). AC: Acoustically measured DTFs. BASE: Base mesh (pinna: 1-mm edge length; head: 2.5 mm).

TABLE 1. Mesh parameters, computation time required to calculate HRTFs, and the modeled localization performance.

Average Edge Length (mm)		Number of Elements	Computation Time (hours)	Polar Error (°)	Quadrant Error (%)
Pinna	Head				
Acoustic Measurement		-	-	29	9.3
0.4	0.5	2 488 720	-	-	-
1	2.5	111 456	19	30	11
2	2	87 706	14	32	12
3	3	44 190	8	33	14
4	4	25 442	6	38	17
5	5	16 464	4.5	40	17

meshes (Fig. 3c). Also, the modeled response probabilities (Fig. 3d) suggest a more blurry localization with larger edge lengths, an effect further revealed by the localization performance parameters (Table 1). Generally, the localization performance monotonously decreased with increasing average edge length, showing indications for the required average edge length in future HRTF calculations.

CONCLUSIONS

In this study, we investigated a method for capturing the geometry of the human head and pinna, and for calculating HRTFs, aiming at satisfying requirements for perceptually valid HRTFs. The BEM was coupled with the fast-multipole method to calculate free-field DTFs for frequencies up to 20 kHz. With the smallest average edge length computed, measured and calculated DTFs were similar. The results of a sagittal-plane localization model suggest a similar localization performance with the measured and calculated DTFs. Further results on meshes with a larger average edge length indicate that the requirements on the mesh quality can be lowered.

- 1 <http://www.cecem.de/gehoerschutz.html>, last viewed on 18.1.2013
- 2 <http://www.ge-mcs.com/en/radiography-x-ray/ct-computed-tomography.html>, last viewed on 18.1.2013
- 3 <http://www.volumegraphics.com>, last viewed on 18.1.2013
- 4 <http://meshlab.sourceforge.net>, last viewed on 18.1.2013
- 5 <http://www.geomagic.com>, last viewed on 18.1.2013
- 6 http://www.zcorp.com/documents/491_8749-ZScanner700_CX_SpecSheet.pdf, last viewed on 18.1.2013
- 7 <http://3d-coat.com/>, last viewed on 18.1.2013
- 8 <http://www.openflipper.org>, last viewed on 18.1.2013
- 9 <http://sourceforge.net/projects/amttoolbox/>, last viewed on 18.1.2013

ACKNOWLEDGMENTS

We thank Christa Rothner for the cap design, Florian Rist (Institute of Arts and Design, Vienna University of Technology) for the ear impressions, Daniel Habe (Österreichisches Gießerei-Institut, Leoben) for the CT Scans, and Matthias Kucera (Ludwig Boltzmann Institute for Archaeological Prospection and Virtual Archaeology, Vienna) for the head scans. We are also grateful to our subjects for their patience during the longsome measurements. This study was supported by the Austrian Science Fund (FWF, P 24124-N13).

REFERENCES

- Baumgartner, R., Majdak, P., and Laback, B. (2013). "Assessment of Sagittal-Plane Sound Localization Performance in Spatial-Audio Applications," In J. Blauert and A. Kohlrausch (Eds.), *The Technology of Binaural Listening* (Springer, NY, USA), Chapter 4.
- Blauert, J. (1997). *Spatial hearing: The Psychophysics of Human Sound Localization* (MIT-Press, Cambridge, MA), 494 pages.
- Burton, A. J., and Miller, G. F. (1971). "The application of integral equation methods to the solution of exterior boundary-value problems," *Proc. R. Soc. Lond. A* **323**, 201–210.
- Gumerov, N. A., and Duraiswami, R. (2009). "A broadband fast multipole accelerated boundary element method for the three dimensional Helmholtz equation," *The Journal of the Acoustical Society of America* **125**, 191–205.
- Kahana, Y., and Nelson, P. A. (2006). "Numerical modelling of the spatial acoustic response of the human pinna," *Journal of Sound and Vibration* **292**, 148–178.
- Kahana, Y., and Nelson, P. A. (2007). "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models," *Journal of Sound and Vibration* **300**, 552–579.
- Katz, B. F. G. (2001). "Boundary element method calculation of individual head-related transfer function I Rigid model calculation," *The Journal of the Acoustical Society of America* **110**, 2440–2448.
- Katz, B. F. G. (2001). "Boundary element method calculation of individual head-related transfer function II Impedance effects and comparisons to real measurements," *The Journal of the Acoustical Society of America* **110**, 2449–2455.
- Kreuzer, W., Majdak, P., and Chen, Z. (2009). "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," *The Journal of the Acoustical Society of America* **126**, 1280–1290.
- Lorenson, W. E., and Cline, H. E. (1987). "Marching cubes: A high resolution 3D surface construction algorithm," *SIGGRAPH Comput. Graph.* **21**, 163–169.
- Majdak, P., Balazs, P., and Laback, B. (2007). "Multiple exponential sweep method for fast measurement of head-related transfer functions," *Journal of the Audio Engineering Society* **55**, 623–637.
- Marburg, S. (2002). "Six boundary elements per wavelength Is that enough?," *Journal of Computational Acoustics* **10**, 25–51.
- Middlebrooks, J. C. (1999). "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *The Journal of the Acoustical Society of America* **106**, 1493–1510.
- Middlebrooks, J. C., and Green, D. M. (1990). "Directional dependence of interaural envelope delays," *The Journal of the Acoustical Society of America* **87**, 2149–2162.
- Möbius, J., and Kobbelt, L. (2012). "OpenFlipper: an open source geometry processing and rendering framework," *Proceedings of the 7th international conference on Curves and Surfaces* (Springer-Verlag, Berlin, Heidelberg), pp. 488–500.
- Møller, H. (1992). "Fundamentals of binaural technology," *Applied Acoustics* **36**, 171–218.
- Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (1995). "Head-related transfer functions of human subjects," *Journal of the Audio Engineering Society* **43**, 300–321.
- Morse, P. M., and Ingard, K. U. (1987). *Theoretical Acoustics* (Princeton University Press), 949 pages.
- Oppenheim, A. V., Schaffer, R. W., and Buck, J. R. (1999). *Discrete-time signal processing* (Prentice Hall), 904 pages.
- Søndergaard, P., and Majdak, P. (2013). "The auditory-modeling toolbox," In J. Blauert (Ed.), *The technology of binaural listening* (Springer, Berlin–Heidelberg–New York NY), Chapter 2.
- Treeby, B. E., Pan, J., and Paurobally, R. M. (2007). "An experimental study of the acoustic impedance characteristics of human hair," *Journal of the Acoustical Society of America* **122**, 2107–17.
- Treeby, B. E., Pan, J., and Paurobally, R. M. (2007). "The effect of hair on auditory localization cues," *The Journal of the Acoustical Society of America* **122**, 3586–3597.
- Treeby, B. E., Paurobally, R. M., and Pan, J. (2007). "The effect of impedance on interaural azimuth cues derived from a spherical head model," *The Journal of the Acoustical Society of America* **121**, 2217–2226.

Chapter 5

Numerical calculation of listener-specific HRTFs and sound localization: Microphone model and mesh discretization

This work was published as:

Ziegelwanger, H., Majdak, P., and Kreuzer, W. (2015). "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization," *The Journal of the Acoustical Society of America*, **138**, 208-222. [doi:10.1121/1.4922518](https://doi.org/10.1121/1.4922518)

The initial idea and design of the study was collaborative work of all authors. I, as the first author, evaluated the geometric details of the head and pinnae models, numerically calculated and acoustically measured HRTFs, analyzed the HRTFs based on model predictions, designed and conducted the sound-localization experiments, analyzed the results, generated all figures and wrote the initial draft of the manuscript. The co-authors provided feedback at every stage of the study, particularly in the interpretation of the results, and they helped writing the manuscript.

Co-author 1:

Co-author 2:

Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization

Harald Ziegelwanger,^{a)} Piotr Majdak, and Wolfgang Kreuzer

Acoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, A-1040 Vienna, Austria

(Received 20 September 2014; revised 22 May 2015; accepted 3 June 2015; published online 10 July 2015)

Head-related transfer functions (HRTFs) can be numerically calculated by applying the boundary element method on the geometry of a listener's head and pinnae. The calculation results are defined by geometrical, numerical, and acoustical parameters like the microphone used in acoustic measurements. The scope of this study was to estimate requirements on the *size* and *position* of the microphone model and on the *discretization* of the boundary geometry as triangular polygon mesh for accurate sound localization. The evaluation involved the analysis of localization errors predicted by a sagittal-plane localization model, the comparison of equivalent head radii estimated by a time-of-arrival model, and the analysis of actual localization errors obtained in a sound-localization experiment. While the average edge length (AEL) of the mesh had a negligible effect on localization performance in the lateral dimension, the localization performance in sagittal planes, however, degraded for larger AELs with the geometrical error as dominant factor. A microphone position at an arbitrary position at the entrance of the ear canal, a microphone size of 1 mm radius, and a mesh with 1 mm AEL yielded a localization performance similar to or better than observed with acoustically measured HRTFs. © 2015 Acoustical Society of America.

[<http://dx.doi.org/10.1121/1.4922518>]

[ELP]

Pages: 208–222

I. INTRODUCTION

Head-related transfer functions (HRTFs) describe the direction-dependent filtering of the incoming sound at the entrance of the ear canal due to sound reflection, shadowing, and diffraction caused by the head, torso, and the pinnae (Møller *et al.*, 1995; Wightman and Kistler, 1989). HRTFs contain listener-specific directional acoustic features, which are analyzed by the auditory system for sound-source localization. Particularly, two different types of features are relevant for the localization process. First, interaural differences, like interaural time differences (ITDs) and interaural level differences (ILDs), are relevant for localization in the lateral dimension (Macpherson and Middlebrooks, 2002). These differences are mostly defined by a listener's head dimensions due to sound scattering and shadowing (Algazi *et al.*, 2001b; Kuhn, 1977; Ziegelwanger and Majdak, 2014). Second, spectral features are relevant for vertical localization and front-back discrimination of sound-source positions. These spectral features are mostly characterized by the fine geometrical details of a listener's pinnae (Bronkhorst, 1995; Hebrank and Wright, 2005; Middlebrooks, 1999a).

HRTFs are usually acquired in an acoustical measurement procedure (Møller *et al.*, 1995). Acoustically measured HRTFs yield localization performance similar to that obtained in free-field listening (Bronkhorst, 1995; Middlebrooks, 1999a). However, the acoustic measurement is a difficult and resource-demanding procedure. The measurement procedure

might be uncomfortable for the listener because of the insertion of small in-ear-microphones and the requirement to sit still during the whole measurement.

HRTFs can also be calculated based on physical and geometrical models. While formerly, various physical models were developed to calculate temporal and spectral features analytically for simple geometrical objects and shapes (Kuhn, 1977; Lopez-Poveda and Meddis, 1996), in recent years, various techniques were used to numerically calculate HRTFs based on the head geometry and the complex shape of the pinna. HRTFs were numerically calculated using the *boundary element method (BEM)* (Katz, 2001a), the *ultra weak variational formulation* (Huttunen *et al.*, 2007), and the *finite-difference time-domain method* (Takemoto *et al.*, 2012). In general, these techniques aim at solving an acoustic scattering problem by numerically solving the Helmholtz equation. In the numerical calculation of HRTFs the scatterer is defined by a listener's geometry. Listeners' geometries were captured by laser scanning (Katz, 2001a; Kreuzer *et al.*, 2009), by magnetic resonance imaging (MRI, Jin *et al.*, 2014), by computerized tomography (CT, Mey *et al.*, 2008), and by photogrammetric reconstruction (Rébillat *et al.*, 2014; Reichinger *et al.*, 2013). When the listener's geometry is known, the numerical HRTF calculation is done by imitating the acoustic measurement procedure; a *virtual* sound source is placed at different positions around the discretized representation of the head and pinnae geometry, i.e., the three-dimensional (3D) polygon mesh, and the HRTFs are obtained from the ratio of the calculated sound pressure at the ear canal and the calculated sound pressure at the origin of the head without the presence of the head.

^{a)}Electronic mail: harald.ziegelwanger@oeaw.ac.at

Much research effort has been done on the numerical calculation of HRTFs by the use of the BEM (Gumerov *et al.*, 2010; Kahana and Nelson, 2006; Katz, 2001a,b; Kreuzer *et al.*, 2009). The computational effort of the BEM increases with increasing frequency and, thus, the frequency range is limited by computer memory and computational power. While the first calculations (Katz, 2001a,b) were limited to frequencies below 6 kHz, the application of the multi-level fast-multipole method (ML-FMM) made the BEM feasible for the entire audible frequency range (Gumerov *et al.*, 2010; Kreuzer *et al.*, 2009).

HRTFs were numerically calculated for animals (Mey *et al.*, 2008; Rébillat *et al.*, 2014), for artificial heads (Greff and Katz, 2007; Gumerov *et al.*, 2010; Kahana and Nelson, 2006) and for human heads (Jin *et al.*, 2014; Katz, 2001a; Kreuzer *et al.*, 2009). For animal heads like those of a bat and for artificial heads like those of the KEMAR mannequin, the calculated HRTFs showed a good correspondence to acoustically measured HRTFs (Gumerov *et al.*, 2010; Mey *et al.*, 2008). Results for human heads are twofold. HRTFs calculated for meshes obtained from laser scans (Katz 2001a; Kreuzer *et al.*, 2009) showed obvious visible differences in the amplitude spectra when compared to measured HRTFs. HRTFs calculated for meshes obtained by MRI showed spectral correlation coefficients above 0.72 when compared to acoustically measured HRTFs (Jin *et al.*, 2014).

In general, quantifying the similarity of different HRTF sets and, thus, validating numerically calculated HRTFs is challenging. HRTFs can differ between measurements even for the same listener (Majdak *et al.*, 2013) or artificial heads (Andreopoulou *et al.*, 2015), thus, differences between the amplitude spectra of acoustically measured and numerically calculated HRTFs are unavoidable. While potential reasons for such differences can be the position of the microphone (Greff and Katz, 2007), the perceptual relevance, however, is unclear as a comparison of HRTF amplitude spectra, even when performed on a logarithmic frequency scale, is not well correlated with human perception (Middlebrooks, 1999a). Acoustically measured HRTFs were validated in sound-localization experiments (Bronkhorst, 1995; Middlebrooks, 1999a) by comparing listener-specific localization-performance parameters for real and virtual sound sources. Recently, various models have been developed to estimate such parameters (Baumgartner *et al.*, 2013; Ziegelwanger and Majdak, 2014) and they can be used to quantify the similarity of HRTF sets on a perceptual level (Ziegelwanger *et al.*, 2013).

While acoustically measured HRTFs can be affected by measurement noise, room reflections, and post-processing artifacts, results of the numerical HRTF calculation depend on geometrical, numerical, and acoustical parameters (Treeby and Pan, 2009). First, numerically calculated HRTFs depend on the accuracy and resolution of the geometric representation of the listener's head and pinnae (Kreuzer *et al.*, 2009). Problems originate from inaccuracies of the geometry acquisition method and of the mesh processing that is required for the numerical calculations. We refer to these problems as *geometrical errors*. Second, if the sound field on the boundary is not adequately approximated then the HRTF calculation suffers from large *numerical errors*

(Marburg, 2002). Note that both geometrical and numerical errors are generally unavoidable because of the discretization of the head and pinnae surface as a 3D polygon mesh. Third, the parameters like the *virtual microphone*, i.e., the model simulating position and size of the microphone used in acoustic HRTF measurements (Greff and Katz, 2007; Kreuzer *et al.*, 2009; Mey *et al.*, 2008), and boundary conditions, e.g., admittance of skin and hair (Treeby *et al.*, 2007a), may result in *acoustical errors*. While results from Katz (2001b) and Treeby *et al.* (2007b) indicate a negligible effect of the boundary admittance on amplitude spectra in the median plane, requirements on the remaining parameters for numerical HRTF calculation and their impact on sound-localization performance, especially on *vertical* sound-localization performance, have not been clarified yet.

The aim of this study was to systematically address these issues, to separate the effects of the different error types, and to identify the relevant parameters required for *perceptually* valid numerically calculated HRTFs in terms of a good sound-localization performance. The effect of the virtual microphone and the effect of the mesh resolution were systematically evaluated. The separation of the different error types was not trivial. On the one hand, position and size of the virtual microphone model parameters are limited by the resolution of the mesh. On the other hand, the results of the numerical HRTF calculation, when investigating the effect of geometrical and numerical errors, depend on the chosen parameters for the virtual microphone model. Thus, first the requirements for the virtual microphone model were evaluated for the best available mesh resolution and accuracy. Then, effects of geometrical and numerical errors were evaluated for a virtual microphone fulfilling these requirements. Both effects were evaluated using model predictions. For the most relevant conditions, numerically calculated HRTFs were finally verified in a sound-localization experiment.

II. GENERAL METHODS

A. Subjects

Three subjects participated in this study (NH5, NH130, and NH131 from the ARI database²). The subjects had different pinna sizes covering a wide range of sizes usually found in human listeners. Figure 1 shows the geometry of left and right ears of all subjects (for the geometry acquisition see Sec. IID). Table I shows the corresponding anthropometric data³ of the subjects together with those of KEMAR (G.R.A.S., Denmark) with normal pinna (DB-60). Note the different pinna sizes: The ears of NH130 are large compared to those of KEMAR and NH5. The ears of NH131 were the smallest and flattest in our study. All subjects were male and had different haircuts, particularly NH5 had an almost bald head, NH130 had shoulder long hair, and NH131 had a short haircut. The subjects had absolute hearing thresholds within a 20-dB range of the average normal-hearing population in the frequency range between 0.125 and 12.5 kHz.

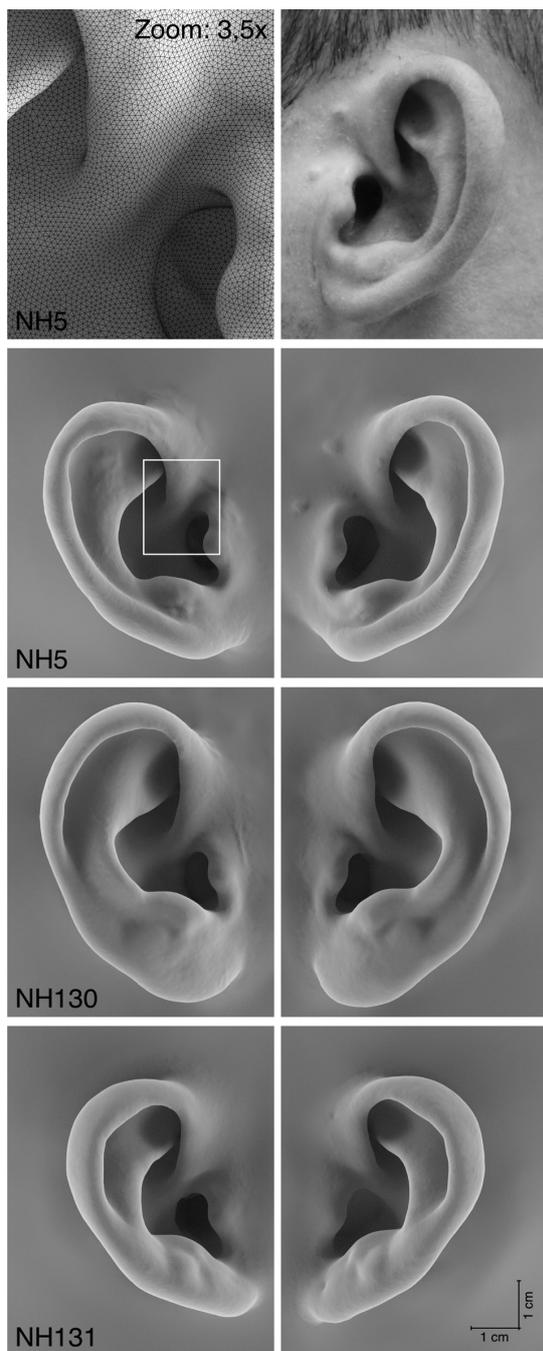


FIG. 1. Mesh renderings of the left (right panels) and right (left panels) ears of the subjects. Zoomed rendering with rendered edges for NH5 (top-left panel) and a photography of NH5's left ear (top-right panel).

B. Acoustic HRTF measurement

Acoustic HRTF measurement was done in a semi-anechoic chamber (Majdak *et al.*, 2010). HRTFs for 1550 directions were measured by placing in-ear-microphones in the subject's ear canals, i.e., blocked ear-canal method (Møller *et al.*, 1995). The horizontal and vertical range was

TABLE I. Anthropometric measurements (x_1 : head width, x_2 : head height, x_3 : head depth, d_5 : pinna height, d_6 : pinna width) of the three subjects and the KEMAR mannequin.

Subject ear	NH5		NH130		NH131		KEMAR
	l	r	l	r	l	r	
x_1 (mm)	156		151		160		140
x_2 (mm)	233		230		240		221
x_3 (mm)	207		195		212		200
d_5 (mm)	61	60	66.9	65.4	56.2	57.1	55.4
d_6 (mm)	27.9	29.1	29.5	31.3	29.5	27.7	27.2

360° and -30° to 80° , respectively. The horizontal resolution was 2.5° and 5° for directions inside and outside of $\pm 45^\circ$, respectively. The vertical resolution was 5° . The measured transfer functions were equalized with the equipment's transfer functions by spectral division. Then, directional transfer functions (DTFs) were calculated (Middlebrooks, 1999b), i.e., the direction independent part of the HRTF was logarithmically subtracted from the HRTFs. The DTFs were temporally windowed with an asymmetric Tukey window to a length of 256 samples (5.33 ms at a sampling rate of 48 kHz).

C. Numerical HRTF calculation

HRTFs were calculated with the BEM using the collocation method with constant elements coupled with the ML-FMM (Chen *et al.*, 2008; Kreuzer *et al.*, 2009).⁴ In general, the BEM numerically solves the boundary integral formulation of the Helmholtz equation, i.e., the boundary integral equation (BIE), in the 3D space (Sauter and Schwab, 2010). The BEM requires a 3D polygon mesh, i.e., a discrete representation of the boundary surface geometry which was the head and pinna in our case. A polygon mesh is usually defined by a set of nodes and a set of triangular or quadrilateral planar elements with these nodes as vertices. In our calculations triangular elements were used.

The sound pressure at the boundary surface is calculated by solving the BIE for a particular sound-source configuration, i.e., an incoming plane wave, a point source, a set of point sources, or radiating mesh elements. In our study, the calculation would take a considerable amount of time because many sound-source configurations had to be considered. In order to keep the computational effort feasible, two different approaches were used to simulate the sound-source configurations.

In the *direct* approach, a point source representing a virtual loudspeaker is placed at an arbitrary position outside the head, the sound pressure on the whole boundary surface is calculated, and the sound pressure at the receiver area, i.e., a mesh element or a set of mesh elements, representing the virtual microphone is evaluated. The BIE has to be solved separately for each virtual loudspeaker position, rendering the direct approach as inefficient for the simulation of many loudspeaker positions. However, many virtual microphone models can be evaluated by averaging the sound pressure on contiguous mesh elements.

In the *reciprocal* approach, the roles of the loudspeaker and microphone are interchanged (Morse and Ingard, 1986) by defining the virtual microphone as a radiating element at the blocked ear canal and evaluating the sound pressure at various positions around the head representing the virtual loudspeakers (Jin *et al.*, 2014; Kreuzer *et al.*, 2009). In contrast to the direct approach, only a single parameter setting for the virtual microphone can be evaluated at once, rendering the reciprocal approach as inefficient for the simulation of many microphone-parameter settings. However, many virtual loudspeaker positions can be evaluated at once.

HRTFs were calculated for 200 frequencies, which were linearly spaced between 0.1 and 20 kHz. The admittance of the boundary surface was set to zero, i.e., the head and pinnae were modeled fully reflective. The resulting HRTFs were inverse Fourier-transformed in order to obtain the head-related impulse responses (HRIRs), which were further re-sampled to a sampling rate of 48 kHz. Finally, DTFs were calculated following the same procedure as for the acoustically measured HRTFs.

In total, HRTFs for 78 conditions were numerically calculated (see details later). All calculations were done on a Linux cluster consisting of eight machines with Intel i7-3820 processors running at 3.6 GHz and 64 gigabyte of RAM each. Given the duration of a single calculation in the order of several hours and the huge number of calculations required, we limited the number of mesh elements to approximately 100 000 in order to finish the calculations in reasonable time. The calculation time for one particular condition was in the order of hours. In total, the duration of the pure calculation process for all conditions in this study was in the order of weeks.

D. Geometry

In the process of geometry acquisition for HRTF calculations the fine details of the pinna geometry, particularly its folds, undercuts, and deep cavities, were of special interest. However, these geometrical details posed technical difficulties during the geometry acquisition process. For instance, some parts of some ears are not always visible from the outside, and those details were difficult to access by surface scanning methods like laser scanners. In contrast, the shape of human head is usually smooth and easy to scan. Thus, two different methods were used for the head and the pinna.

The head geometry was captured by a laser scanner (ZScanner 700CX,⁵ ZCorp). A built-in optical tracking system of this device ensured a globally consistent and accurate scan, even when the subject moved the head during the scanning process. However, the laser scanner was not able to scan hairy surfaces. Thus, the subjects wore a custom tailored cap with holes for the ears (Ziegelwanger *et al.*, 2013). In a post-processing procedure, a 3D triangular polygon mesh was extracted from a volumetric representation with approximately 0.5 mm voxel size. In the final mesh the neck and the ear canal were cut and closed. The sharp cutting edge at the neck was smoothed to avoid edge effects in the BEM.

The pinna geometry was captured using a high-energy industrial computer-tomography scanner (v|tome|x c,⁶ GE Measurement & Control). Such a scanner is not limited in the amount of radiation by medical concerns and is capable to deliver a detailed representation of the pinna geometry. To prevent the subjects from the harmful x-ray radiation, silicone impressions of subjects' pinnae and surrounding area were made (viscosity: 2200 MPa·s, shore hardness: A 22, for more details on the silicon impressions see Ziegelwanger *et al.*, 2013). The impressions were then scanned using 1000 x-ray images spaced in 0.36° intervals (tube settings: 120 kV, 260 μ A, average of three images with 333-ms exposure time each). The resulting volumetric mesh consisted of approximately $2000 \times 2000 \times 650$ voxels and the voxel size was approximately 50 μ m. With the rotational resolution of the x-ray scanner, the accuracy at the outer edge of the largest pinna was 0.21 mm. For all other regions on the pinna, the accuracy of the x-ray scanner was smaller than 0.21 mm. The iso-surfaces separating air and silicone were extracted using automatic thresholding, i.e., a proprietary version of the marching cubes algorithm (Lorenson and Cline, 1987). The effective accuracy of this algorithm was in the range of two to four voxels, i.e., 0.1–0.2 mm.

The pinnae meshes from the CT scans were manually aligned with the pinnae meshes from the laser scans via tie points (Manual Registration, Geomagic⁷). Then the alignment was automatically refined (Best Fit Alignment, Geomagic). The pinnae meshes from the laser scan were removed and the seam-region, i.e., the surrounding area of the pinnae on the head surface, was re-sculpted, smoothed (Voxel-Sculpting, 3D-Coati) in a resolution of below 0.5 mm voxel-size, and water-tightened. Note that the pinnae meshes from the CT scans were not manipulated in this procedure. The final reference meshes consisted of approximately 2.5×10^6 elements and had an average edge length (AEL) and a maximum edge length of 0.47 and 0.77 mm, respectively (for more details on the geometry acquisition, see Reichinger *et al.*, 2013; Ziegelwanger *et al.*, 2013).

The meshes were aligned to match the coordinate system of the loudspeaker grid used in the acoustic HRTF measurement setup. The *auricular points*, i.e., the craniometric points at the center of the ear canal, were placed on the y axis with the left auricular point on the positive half-axis. The midpoint between the auricular points was aligned with the origin of the coordinate system. The match of the head rotation about the interaural axis, i.e., the *interaural rotation*, was more problematic because distinct markers indicating the exact rotation of the interaural horizontal plane, i.e., the horizontal plane with elevation angle of zero, were not available.⁸ This problem has been recently addressed by rotating the coordinate system after the HRTF calculations such that the correlation between the acoustically measured and numerically calculated HRTFs was maximized (Jin *et al.*, 2014). We assumed that a mismatch between the interaural horizontal plane of the acoustic measurements and the numeric calculations would result in a simple polar bias in the localization responses, which can be *post hoc* corrected (see the discussion later). Thus, we approximately aligned the interaural rotation to match photographs from the

acoustic measurement, where the nasal point, i.e., the tip of the nose, seemed to point towards the interaural horizontal plane. In our meshes, the nasal point was placed on the positive half-axis of the x axis and the plane connecting the auricular points and the nasal point defined the interaural horizontal plane.

The reference meshes resulted in a detailed representation of the geometry (AEL < 0.5 mm), however, they had too many elements for feasible numerical calculations (more than 2.5×10^6 elements). For accurate BEM results, the length of edges in the mesh should be a fraction of the smallest simulated wavelength. In an empirical study on simple geometric objects, Marburg (2002) showed that the relative numerical error of the BEM is below 15% when the edge length is a sixth of the smallest simulated wavelength. We refer to the corresponding frequency as the critical frequency, f_{crit} . For a critical frequency of 20 kHz, the criterion from Marburg (2002) leads to an edge length of 2.7 mm. However, a simple increase of the AEL from 0.5 to 2.7 mm resulted in clearly visible geometrical deformations of the pinna. Since our goal was to systematically investigate the AEL required for perceptually valid HRTFs, the finest mesh in our conditions had to include all pinna details and to accurately represent the geometry.

In order to determine the largest possible AEL still accurately representing the geometrical details of a pinna, we used the Hausdorff distance (Cignoni *et al.*, 1998) implemented in Meshlab.⁹ The Hausdorff distance is a metric describing the deviation of parts of two meshes. We calculated the Hausdorff distance between the reference geometry and re-meshed geometries with AELs ranging from 0.5 to 20 mm. Figures 2(c), 2(d), and 2(e) show the left ear of NH5 with the Hausdorff distance for meshes with AELs of 1, 2.7, and 4 mm, respectively (darker colors indicate larger Hausdorff distances). Figure 2(a) shows the maximum Hausdorff distance evaluated for all meshes of all subjects as a function of the AEL. The solid line represents a sigmoid fit to these data. For the AELs of 0.5 and 1 mm, the maximum Hausdorff distances were in the range of 0.5 mm. Beyond the AEL of 2 mm the maximum Hausdorff distance increased rapidly, with a distance of 2.25 mm for the AEL of 2.7 mm. Note that distances larger than the AEL indicate an

inaccurate geometry representation due to the discretization as compared to the reference mesh. In order to find the largest AEL still representing the geometry accurately, we evaluated the normalized Hausdorff distance, i.e., the maximum Hausdorff distance divided by the corresponding AEL. Figure 2(b) shows the normalized Hausdorff distance as a function of the AEL, showing a clear minimum around the AEL of 1 mm. Thus, when compared to the reference geometry, it seems like meshes with an AEL of 1 mm would still accurately represent the pinna geometry and can be used as the finest meshes for further calculations.

Based on these findings, a mesh was created for each subject. The meshes had different resolutions inside and outside the pinna region, similar to the method used in Katz (2001a). The AEL was 1 mm in the pinna region and 2.5 mm at the rest of the head. This 1-mm mesh satisfied both, the global numerical recommendation of six elements per wavelength on the head and the local geometrical accuracy requirement of 1 mm AEL on the pinnae. This mesh was obtained by applying three processing steps. First, the whole reference mesh was re-meshed to an AEL of 1 mm. Then, everything except the pinna region was re-meshed to an AEL of 2.5 mm with a transition region of four polygon rows and an AEL of 1.5 mm. Finally, the transition region was additionally relaxed to reduce degenerated non-equilateral triangles at the region seams. The 1-mm meshes consisted of approximately 110 000 elements and had the critical frequency of 24 kHz.

As a rough check for the methods described so far, we calculated HRTFs following the methods from Kreuzer *et al.* (2009). Spatial correlation coefficients, i.e., the correlation of acoustically measured and numerically calculated HRTFs over all directions as a function of the frequency, were evaluated as described in Jin *et al.* (2014). Figure 3 shows the spatial correlation coefficient for the left and the right pinnae of our subjects and gray areas indicating two ranges of the coefficients observed by Jin *et al.* (2014). The light gray area shows the range of the coefficients of all 61 subjects from Jin *et al.* (2014). The dark gray area shows the range given by the 20 listeners showing the best matches from Jin *et al.* (2014). The spatial correlation coefficients of our three subjects seem to be well-covered by the range from Jin *et al.*

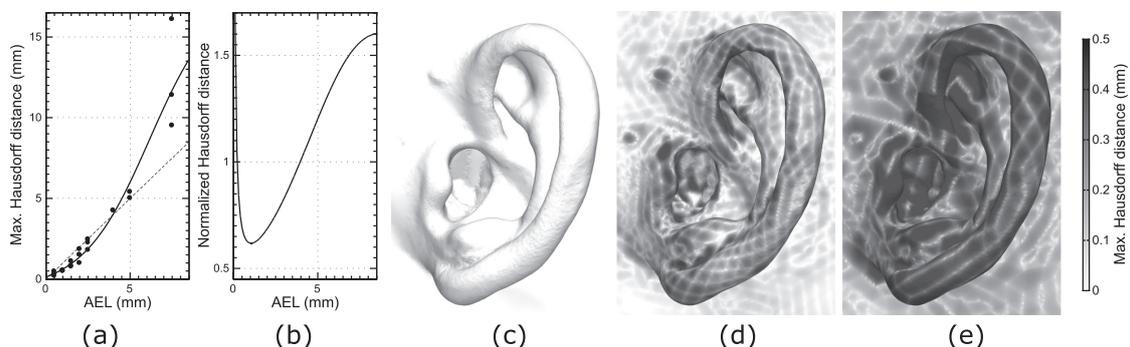


FIG. 2. Maximum Hausdorff distance (a) and normalized Hausdorff distance (b) as a function of the mesh resolution described by the AEL (dots: individual data for three subjects; line: sigmoid fit). Visualization of the Hausdorff distance for an AEL of 1 (c), 2.8 (d), and 4 mm (e) as a function of the location on the left reference ear of NH5. The color bar is limited to 0.5 mm.

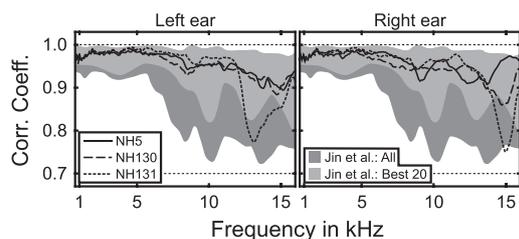


FIG. 3. Spatial correlation coefficient between the acoustically measured and numerically calculated HRTFs for left and right ears of NH5, NH130, and NH131. The gray areas show the range of correlation coefficients observed in Jin *et al.* (2014), for all their subjects (light gray) and for their 20 best subjects (dark gray).

(2014) indicating promising results on the level of spectral comparison.

III. EFFECT OF THE VIRTUAL MICROPHONE

Different microphone configurations have been used to acoustically measure HRTFs, e.g., probe microphones were used for creating the CIPIC database (Algazi *et al.*, 2001c), the blocked-ear-canal method in combination with probe microphones was used in Møller *et al.* (1995), and the blocked-ear-canal method with electret microphones was used in Majdak *et al.* (2010). In the numerical calculation of HRTFs, these microphones are virtually represented by a model, i.e., a virtual microphone (Jin *et al.*, 2014; Kreuzer *et al.*, 2009). Interestingly, the particular position of the virtual microphone has an effect on the amplitude spectra of calculated HRTFs because of interferences and, thus, on features like peaks and notches (Kreuzer *et al.*, 2009). These spectral features are mostly relevant for sagittal-plane sound-source localization, thus, the parameters of the virtual microphone may have a perceptual impact. On the one hand, one might argue that the position and the size of the actual microphone must be also represented in the numeric calculations. On the other hand, the HRTFs from various HRTF databases, all using different microphones, seem to be able to provide similar localization performance. Thus, it might be that the impact of the microphone is perceptually not significant. Hence, in this section, we aim at estimating the requirements on the virtual microphone in numeric calculations with respect to sound localization.

Acoustically measured HRTFs have been validated in terms of listener-specific localization performance, e.g., HRTFs measured with probe microphones were validated in Bronkhorst (1995) and HRTFs measured with electret microphones were validated in Majdak *et al.* (2010). Thus, in this study the acoustically measured HRTFs were assumed to correspond to the internal template of HRTFs a listener is calibrated to Langendijk and Bronkhorst (2002). Based on this assumption, the potential perceptual effect of the virtual microphone, namely, its position and size, on the numerically calculated HRTFs was evaluated. Since the acquisition of localization performance in psychoacoustic experiments would be too time intensive for such a large number of conditions, microphone parameters were evaluated by analyzing sound-localization performance predictions estimated by a

sagittal-plane sound-localization model (Baumgartner *et al.*, 2013).

We assumed that the useful positions of the virtual microphone can be reduced to the surface surrounding the ear canal, i.e., the concha. HRTFs are usually measured for hundreds of source positions and the concha consists of approximately 2000 elements (in the 1-mm mesh). Since each mesh element represents a potential virtual microphone position, theoretically, hundreds of thousand of numerically calculated HRTFs would be needed. In order to keep the computational effort small, we used the direct approach, in which for a single sound-source position, the pressure at the whole boundary was calculated. Then, for that sound-source position the effect of the position and size of the virtual microphone model was evaluated by averaging the sound pressures on different contiguous elements. Assuming an impact on sagittal-plane sound-source localization only, the calculations were done for sound-source positions in the median plane.

In Secs. III A–III C, we show detailed analysis for subject NH5 (the same subject as in Kreuzer *et al.*, 2009). Based on the analysis for that listener, we derive quantitative parameters describing the perceptual effect of the virtual microphone, and then show these parameters for all three listeners to support our key findings.

A. Methods

1. Material

The complex-valued head-surface sound pressure for the 1-mm meshes of the three subjects (compare Sec. II C) was calculated. Point sources representing loudspeakers were placed in the median-plane at a distance of 1.2 m with a polar incident angle ranging from -30° to 210° and a resolution of 15° . In total, 17 sound-source positions were used. After the inverse Fourier transform of the complex surface pressure spectrum, we obtained 17 HRIRs for each triangle in the meshes. All HRIRs were re-sampled to a sampling rate of 48 kHz. HRIRs for different microphone sizes were simulated by averaging HRIRs for contiguous sets of triangles. For the numerically calculated and acoustically measured HRIRs, median-plane DTFs were calculated following the procedure as described in Sec. II B with the exception of calculating the direction-independent part, i.e., the common transfer function (Middlebrooks, 1999a), from HRTFs in the median plane only.

2. Localization model

The perceptual similarity of the acoustically measured and numerically calculated DTFs was evaluated by means of the sagittal-plane sound-localization model (Baumgartner *et al.*, 2013). This model compares a given DTF to an internal template, i.e., an internal representation of the DTFs to which a listener's auditory system is thought to be calibrated to. In our case, the internal template was assumed to be represented by the listener-specific acoustically measured DTFs. The model is quite complex, consists of many stages,

and was in detail described in Baumgartner *et al.* (2013) and Majdak *et al.* (2014).

For the calibration, the model requires an uncertainty parameter describing the sensitivity of the listener. The uncertainty parameter is assumed to distinguish between good and poor localizers and can be obtained from sound-localization experiments. We assumed our listeners to be average localizers and we calibrated the model with the uncertainty of 1.9.

The model describes the listener-specific localization performance by means of the *polar root mean square (RMS) error (PE)* and the *quadrant error rate (QE)*. We used definitions exactly as in Middlebrooks (1999a). The QE is a global parameter representing the amount of errors larger than 90° and can be interpreted as confusions between the hemifields (top, down, front, back). The PE is a local parameter representing the accuracy and precision of the localization when the hemifields are not confused (for more details, see Baumgartner *et al.*, 2013).

3. Conditions

The virtual microphone was an approximately circular area described by its position and size.

The size of the virtual microphone was defined by the circle radius, which was increased from 1 mm, i.e., a single mesh element, to 6 mm by incrementally adding rings of neighboring elements. Figure 4 shows four exemplary microphone positions A, B, C, and D and four microphone sizes for position C, for the left-ear mesh of NH5 zoomed to the region of the concha. A, B, C, and D represent the positions center of the ear canal, edge of the ear canal, outside the ear canal, and outside the concha, respectively. The 6-mm virtual microphone consisted of a center element encircled by six rings of elements.

The position of the virtual microphone was defined as the midpoint of the element in the case of the 1-mm microphone, and the midpoint of the central element in the cases of larger microphone sizes. The position was varied within a radius of 45 mm around the approximate center of the ear

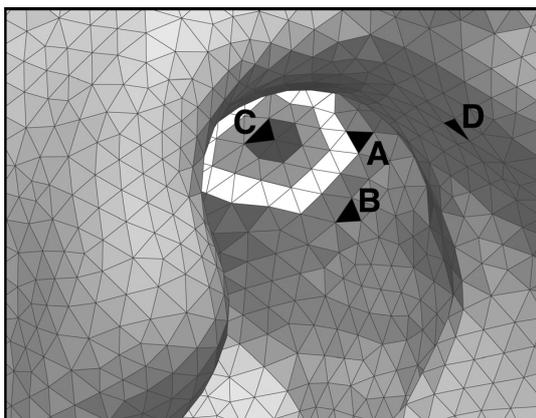


FIG. 4. Geometry of NH5's left concha represented as a triangular polygon mesh (1 mm AEL) with exemplary virtual-microphone positions (A to D) and sizes at position C.

canal. In total, approximately 25 000 conditions for each ear were systematically evaluated.

B. Results and discussion

1. Spectral features

Figure 5 shows DTF amplitude spectra calculated for the microphone positions A, B, and C and a microphone size of approximately 1 mm which corresponds to a single element at our mesh. The lateral and polar angle of the sound source were 0° and 45°, respectively. The DTF amplitude spectrum from the acoustic measurement is shown for comparison. This figure directly addresses Fig. 11 from Kreuzer *et al.* (2009) where a frequency shift of spectral features was observed as the virtual microphone was moved from one element to a neighboring element. In Kreuzer *et al.* (2009), the largest frequency shift occurred for the same sound-source position. In our Fig. 5, a small frequency shift of the notch at 13.5 kHz, the absence of the peak 6 and 7 kHz, and the absence of the sharp notch at 9.5 kHz can be observed. Note that the acoustic measurement was band-limited by approximately 18 kHz.

In order to estimate the impact of the microphone size, DTFs for three microphone sizes at the positions A to D were evaluated. Figure 6 shows DTF amplitude spectra calculated for these microphone positions A to D with the size of the virtual microphone as parameter. The lateral and polar angle of the sound source was 0° and 45°, respectively. The radius of the virtual microphone was approximately 1, 3, and 6 mm. When compared to the acoustically measured DTFs, the DTFs calculated for the larger microphone sizes appear spectrally smeared, even for the position A, which most probably represents the actual microphone position from the acoustic measurement. Further, with larger size, additional spectral features arise. For instance, at the position D when the microphone size increased to 6 mm [Fig. 6(D)], an additional notch occurred at 10.5 kHz and a corresponding notch cannot be found in the acoustically measured DTFs. Such additional features probably result from including the areas

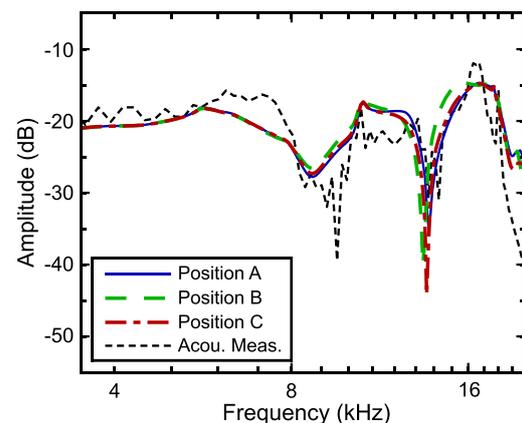


FIG. 5. (Color online) DTF amplitude spectra for virtual-microphone positions A to C (compare Fig. 4) for a single triangle representing the virtual microphone.

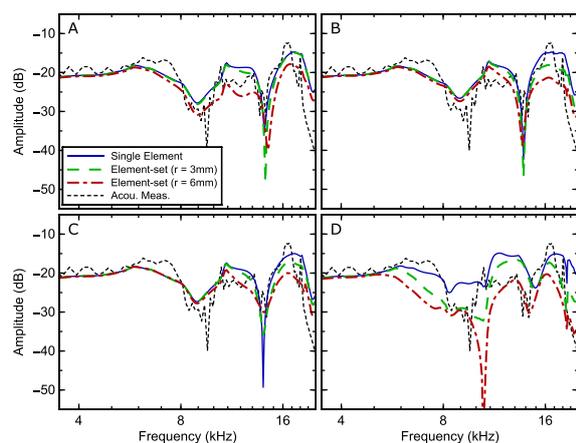


FIG. 6. (Color online) DTF amplitude spectra for virtual-microphone positions A to D (compare Fig. 4) and sizes of 1, 3, and 6-mm radius.

outside of the ear canal in the average of the sound pressure over the microphone area. In summary, the larger sizes of the virtual microphone seem to yield DTFs with smeared features (as at position A, B, and C) and/or features, which are unreasonable in comparison to acoustically measured DTFs (as at position D). The microphone size of approximately 1 mm yielded the smallest spectral difference and, thus, seems to be the most appropriate choice.

As for the size, it seems that multiple elements for the virtual microphone are of no advantage. As for the position, a small shift in position of the virtual microphone placed near the ear canal results in frequency shifts of the spectral features. But do these shifts affect the sound-localization performance?

2. Sound-localization performance

In order to address this question, localization performance was predicted as a function of the microphone position and size. Figure 7(c) shows the left-ear concha of NH5 and two regions (denoted by vertical and horizontal lines) along which the microphone positions were investigated. Figures 7(a) and 7(b) show the relative PEs and QEs for the microphone positions from the vertical region. Figures 7(d) and 7(e) show the relative PEs and QEs for the microphone positions from the horizontal region. In each of the four panels (a, b, d, e), the different lines represent virtual microphone sizes with radii of 1, 3, and 6 mm. The relative PEs and QEs are PEs and QEs relative to references, i.e., the smallest PEs and the QEs in the concha, respectively. These references are shown in Table II, labeled as SIM.

The predicted localization performance was generally within the range which was usually found in sound-localization experiments (Middlebrooks, 1999a). It can be seen in Fig. 7 that the position had a notable effect on the predicted localization performance. Receiver elements within the blocked ear canal resulted in small localization errors. Receiver elements at the border or outside the ear canal resulted in larger localization errors. The effect of the virtual microphone size depended on the position of the

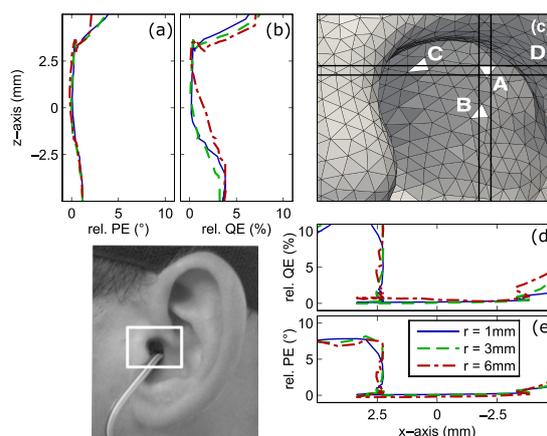


FIG. 7. (Color online) Geometry of NH5's left concha (c) with horizontal and vertical lines representing horizontal and vertical regions, respectively, used for the evaluation of the relative PE [(a) and (e)] and the relative QE [(b) and (d)] for virtual-microphone sizes of 1, 3, and 6-mm radius. The references for the relative parameters are shown in Table II (condition SIM).

receiver elements. The specific microphone positions underline that effect. Positions A and C resulted in a similar predicted localization performance suggesting that a specific position is not required as long as being in the area of the ear canal. Positions B and D resulted in a deteriorated localization performance, confirming the intuitive assumption that a receiver element outside the ear canal is not recommended. Interestingly, the effect of the microphone size was small as long as the receiver element was approximately in the range of the blocked ear canal (position A), in contrast to the larger effect of the microphone size for microphone positions at the border of the ear canal (positions B and D). This indicates that consideration of the actual microphone size is not required in HRTF calculations.

In order to finally evaluate the optimal microphone positions within the ear for all three subjects, median-plane HRTFs were calculated for each single mesh element as the virtual microphone on the pinna, and localization performance was predicted for all six ears, see Fig. 8. The gray-value in the area of each element encodes the relative PEs

TABLE II. Predicted localization performance for acoustically measured HRTFs (AC), numerically calculated HRTFs (SIM), and HRTFs of KEMAR. L, R: left and right ear, respectively.

Subject	HRTFs	PE (°)		QE (%)	
		L	R	L	R
NH5	AC	28.7	29.2	7.9	14.8
	SIM	30.0	29.9	12.8	10.7
	KEMAR	35.1	34.7	24.9	21.2
NH130	AC	24.7	22.8	6.3	9.0
	SIM	25.9	26.6	10.7	9.3
	KEMAR	33.5	35.6	19.6	21.1
NH131	AC	30.7	27.4	13.5	10.0
	SIM	30.4	30.7	11.0	11.5
	KEMAR	38.0	33.3	25.8	18.9

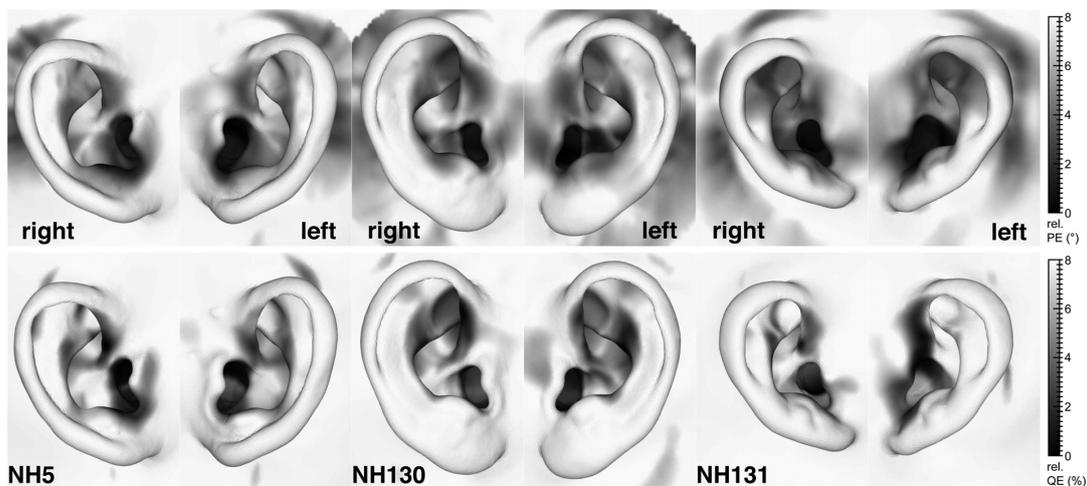


FIG. 8. Relative PEs and relative QEs as function of the virtual-microphone position for a microphone size of 1 mm. The references for the relative parameters are shown in Table II (SIM).

and the QEs obtained for the corresponding ear. The references of the relative PEs and QEs are shown in Table II (SIM), along with the predicted individual localization performance for the listener-specific acoustically measured HRTFs (AC) and for mannequin HRTFs (KEMAR). The predictions for listener-specific HRTFs (SIM and AC) were in the range of localization performance usually observed for acoustically measured HRTFs (Middlebrooks, 1999a). Further, the predicted errors were significantly smaller than those predicted for non-individual KEMAR HRTFs.

For all ears, the localization errors were small when the virtual microphone was placed in the ear canal and increased rapidly when the virtual microphone was moved beyond the ear canal. Especially, the position of the smallest virtual microphone was not crucial. For that 1 mm size, the predicted PEs and QEs, averaged over the six ears, were in the range of the smallest $PE \pm 0.5^\circ$ and $QE \pm 0.5\%$ (see Table II, SIM) when the virtual microphone was placed within a range of 2.5 ± 1.8 mm at the blocked ear canal.

C. Conclusions

We aimed at estimating the requirements for the details of modeling the actual microphone used in the acoustic measurements. We could observe differences in the amplitude spectra for two neighboring receiver elements, similarly to differences found in Kreuzer *et al.* (2009). These differences were, however, not reflected in the predicted sound-localization performance.

All three sizes of the virtual microphone yielded predicted localization performance in the range of that found for acoustically measured HRTFs. This indicates that a detailed model of the microphone from the acoustic measurements is not required. The predicted performance, however, degraded significantly with increasing the distance of the virtual microphone to the ear canal and that dependence became larger with larger microphone sizes.

This indicates that a small receiver element, here with the size of 1 mm, placed approximately in the center of the blocked ear canal can sufficiently model the microphones for numeric HRTF calculations. It even seems that, permitting for a tolerance of $PE \pm 0.5^\circ$ and $QE \pm 0.5\%$, the exact position of the virtual microphone can deviate from the center by as much as 2.5 mm on average.

IV. EFFECT OF THE DISCRETIZATION: MODEL PREDICTIONS

In this section, the requirements on the boundary discretization were investigated. The discretization of the boundary surface by planar triangles may introduce both geometrical and numerical errors. Geometrical errors are caused by under-sampling the geometry and discarding geometric details. Numerical errors are caused by under-sampling the sound field. Both types of error are addressed in this section. The joint effect of geometrical errors and numerical errors was systematically investigated with *coarsened meshes*, i.e., re-meshed reference meshes with larger AELs. The effect of numerical errors was separated from the effect of geometrical errors by decreasing the AEL in the coarsened meshes, here referred to as *refined meshes*. These refined meshes had the same geometrical accuracy as the coarsened meshes but smaller AELs, hence they did not under-sample the sound field. The impact of the two types of errors on sound localization was evaluated in terms of comparing model predictions for HRTFs of the different meshes. Two models were used: The similarity of temporal features was evaluated by comparing equivalent head radii (Ziegelwanger and Majdak, 2014) and the similarity of spectral features was evaluated by comparing predicted sagittal-in Sec. III.

A. Conditions

For each subject, two sets of meshes were created. The first set was composed of the 1-mm mesh and of four

coarsened meshes with AELs of 2, 3, 4, and 5 mm. The coarsened meshes were generated by re-meshing the reference meshes with a re-meshing algorithm (Botsch and Kobbelt, 2004) implemented in *OpenFlipper*.¹⁰ The statistics of the meshes are shown in Table III. The second set consisted of four refined meshes created by subdividing each triangle in the coarsened meshes to four triangles. For the 2- and 3-mm meshes, this subdivision was limited to the pinna region in order to limit the size of the meshes to approximately 100 000 elements.

B. Methods

For all three subjects, HRTFs were numerically calculated using the reciprocal approach. Based on the results from the previous section, the virtual microphone was a single, uniformly vibrating, triangular element, placed at the approximate center of the ear canal. The velocity of the vibrating element was 0.1 mm/s. HRTFs were numerically calculated for the same loudspeaker grid as used in the acoustical measurement (Majdak *et al.*, 2010). HRIRs and DTFs were calculated as described in Sec. II.

The similarity of numerically calculated and acoustically measured HRTFs was evaluated by applying two models. The first model evaluated the similarity of spectral features by predicting listener-specific PEs and QEs. The model and its parameters were the same as in Sec. III A 2. In addition to listener-specific predictions for the three tested listeners, two benchmarks were calculated aiming to frame the predicted performance into adequate context. The benchmark “own” was the statistics of localization performance predicted for 172 listeners of the ARI,¹¹ CIPIC,¹² and LISTEN¹³ databases under the assumption of listening with own ears, i.e., localizing sound sources with their corresponding listener-specific HRTFs. The benchmark “other” was the statistics of localization performance predicted for NH5, NH130, and NH131 under the assumption of localizing with other’s ears, i.e., the PEs and QEs predicted for localizing sound sources with HRTFs of all 171 other listeners from the three databases.

The second model evaluated temporal features, in particular the broadband time-of-arrival (TOA) in HRIRs, which interaural difference is the relevant feature for localization along the lateral dimension (Macpherson and Middlebrooks, 2002). When considering the head as an approximation of a sphere, the TOA depends on the equivalent head radius (Algazi *et al.*, 2001a). The equivalent head

radius was estimated by a geometrical TOA model (Ziegelwanger and Majdak, 2014). In a nutshell, the TOA model estimates broadband delays from an HRTF set of a listener by the use of the minimum-phase cross-correlation method¹⁴ and fits a spherical head model of the listener to them. The radius of the head model represents the equivalent head radius of the listener and defines the spatial ITD function in an HRTF set. The equivalent head radii were estimated for the numerically calculated HRTFs and for the acoustically measured HRTFs. As a rough estimation, we assumed any difference between equivalent head radii larger than 3.55 mm as evidence for differences in the temporal features because 3.55 mm corresponds to a half of the spatial sampling interval of 7.1 mm assuming a speed of sound by 343 m/s and a sampling rate of 48 kHz. Differences smaller than 3.55 mm were interpreted as insignificant.

C. Results and discussion

Figure 9 shows median-plane amplitude spectra for the acoustically measured [Fig. 9(a)] and numerically calculated DTFs for coarsened meshes [Figs. 9(b) to 9(f)] of NH5’s left ear. Note that the acoustic measurement was band-limited between 0.3 and 18 kHz. Further, the numerical calculations

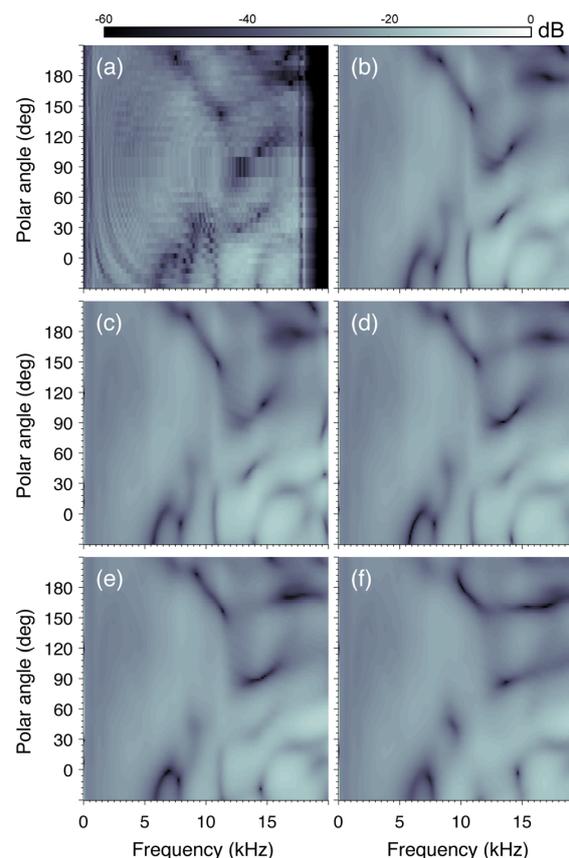


FIG. 9. (Color online) Median-plane DTF amplitude spectra of NH5; acoustically measured (a) and numerically calculated for meshes with an AEL of 1–5 mm [(b)–(f)].

TABLE III. Mesh statistics (standard deviation of edge lengths; minimum and maximum edge length; number of elements averaged over subjects) and critical frequencies.

AEL	Edge length (mm)			Number of elements	f_{crit} (kHz)
	Std.	Min.	Max.		
2	0.2	1.1	3.1	106 526	28.6
3	0.3	1.6	4.0	53 671	19.1
4	0.4	2.3	5.9	28 188	14.3
5	0.5	2.8	7.1	18 143	11.4

did not include a torso model, thus, spectral ripples at low frequencies arising from torso reflections are missing. For the acoustically measured DTFs and the numerically calculated DTFs of the 1-mm mesh, the spectral notches appear to be similar. For larger AELs, these notches change their spectro-spatial position or even disappear, e.g., the notch at 13 kHz and 45° elevation completely disappeared in the 5-mm mesh. For this mesh, spectral features seem to be clearly different compared to those from the acoustic measurement, with differences appearing even at frequencies below 10 kHz.

Figure 10 shows the predicted localization performance in sagittal planes. The gray area shows the standard deviation range of the localization performance observed in Middlebrooks (1999a). For the acoustically measured HRTFs, for the 1- and 2-mm mesh, and for the benchmark “own,” predicted PEs and QEs were within that area. For numerically calculated HRTFs, the predicted performance degraded with increasing AEL. For the 5 mm mesh, the predictions were in the range of localization performance observed for the “other” benchmark. Thus, the range of AELs from 1 mm to 5 mm represents good candidates for further evaluation in actual sound-localization experiments.

For the coarsened meshes, both the PE and QE increased with AEL. A linear regression was fit to the predictions in order to roughly examine the relation between the AEL and predicted localization performance. For the PE, the slope of the linear regression was 1.51°, 1.79°, and 1.92°/mm for NH5, NH130, and NH131, respectively, and 1.74°/mm on average. For the QE, the slope was 0.99%, 2.13%, and 3.19%/mm for NH5, NH130, and NH131, respectively, and 2.10%/mm on average.

Table IV shows the equivalent head radii obtained from the TOA model for all conditions and subjects. For numerically calculated HRTFs, the equivalent head radii were similar to those for the acoustically measured HRTFs indicating that

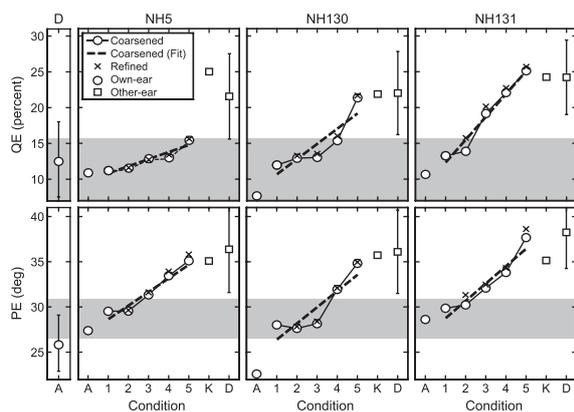


FIG. 10. Localization performance predicted by the sagittal-plane localization model. Mostleft column: Average and ± 1 standard deviation of predictions for all subjects from the databases (ARI, CIPIC, and LISTEN) when listening with own-ear DTFs. Other columns: Predictions for NH5, NH130, and NH131 in conditions with acoustically measured own-ear DTFs (labeled as A), numerically calculated own-ear DTFs for the coarsened and refined meshes (1–5), acoustically measured manikin DTFs of the KEMAR (K), and acoustically measured other-ear DTFs from the databases (D). The gray area shows the average ± 1 standard deviation from Middlebrooks (1999a).

TABLE IV. Equivalent head radii for acoustically measured HRTFs (AC) and numerically calculated HRTFs of coarsened and refined meshes (AEL 1–5 mm).

	Equivalent head radius (mm)					
	Coarsened meshes			Refined meshes		
	NH5	NH130	NH131	NH5	NH130	NH131
AC	91.8	92.4	94.5	—	—	—
1	90.8	89.2	94.6	—	—	—
2	91.4	89.4	94.6	91.4	89.1	92.8
3	91.4	89.4	94.6	91.1	89.3	94.3
4	91.8	90.2	94.5	89.4	89.9	94.3
5	91.1	89.8	95.1	90.7	89.4	94.5

the AEL is not critical for accurate representation of broad-band temporal features in numerically calculated HRTFs.

For refined meshes, the predictions were similar to the predictions for the coarsened meshes. For both PE and QE, the Pearson’s correlation coefficients between the predictions for the coarsened meshes and those for the refined meshes were 0.99, showing further evidence for a very similar performance for both coarsened and refined meshes. Also the equivalent head radii observed for the coarse and refined meshes were similar. This indicates that the performance degradation with increasing AEL was solely based on geometrical errors and the contribution of the numerical errors was negligible. Thus, the AEL effect observed in conditions with coarsened meshes appears to rely on poor discretization of the detailed geometry of the pinna with the consequence that the refined meshes are not required in further HRTF calculations.

V. EFFECT OF THE DISCRETIZATION: LOCALIZATION EXPERIMENTS

In this section, particular conditions from Sec. IV were tested in a sound-localization experiment in order to validate the predictions (and thus the numerical calculations) with respect to sound localization. Seven experimental conditions were tested: listener-specific acoustically measured HRTFs, acoustically measured KEMAR HRTFs, and numerically calculated HRTFs for the 1-mm mesh and the meshes coarsened up to the AEL of 5 mm.

A. Methods

All listeners from the previous sections participated in the sound-localization experiments. The procedure of the sound-localization experiment was based on that from Majdak *et al.* (2010) and identical to that from Majdak *et al.* (2013) with the exception of a small change in the visual virtual environment (see later).

Briefly, the stimuli were Gaussian white noises of 500-ms duration filtered with DTFs corresponding to the tested condition and position. The tested lateral positions ranged from -90° (right) to 90° (left) and the tested elevations ranged from -30° to 210° . The stimuli were presented via headphones (HD 580, Sennheiser) at the level of 50 dB above the individual absolute hearing threshold and level

roving of ± 5 dB was applied. The listeners were immersed in a spherical virtual visual environment. They held a pointer in their hand. The pointer's projection was visualized and recorded as the perceived target position. In contrast to Majdak *et al.* (2013), the head-mounted display was an Oculus Rift (Developer Kit, OculusVR). For each eye, the display provides screens with a field of view of $110^\circ \times 90^\circ$ (horizontal \times vertical dimensions) and a resolution of 800×640 pixels. In the virtual environment, the colors, brightness, and the grid were adapted to the field-of-view such that the immersed listeners were feeling comfortable and relaxed without any signs of dizziness, e.g., the background color was set to light gray, the grid lines were removed and replaced by balls at grid corners. The grid resolution was 45° and 11.25° in the azimuth and elevation, respectively. While balls indicated the horizontal and median plane for orientation were blue, the rest of the balls were black. A red ball indicated the position 0° azimuth and 0° elevation.

Prior to the main experiment, listeners participated in a visual and acoustic training procedure with the goal of using the equipment accurately and of performing stably. In the visual training, the listeners pointed to visual targets within a limited amount of time. They continued the training until they were able to point to 95% of the targets within 4 s with an average angular distance smaller than 2° (for more details see Majdak *et al.*, 2013, p. 2058). In the acoustic training, listeners were localizing sounds filtered with their own DTFs and visual feedback on the actual target position was applied. This training consisted of 300 trials (for more details, see Majdak *et al.*, 2010, p. 462).

In the main experiment, at the beginning of each trial, the listeners were asked to align themselves with the reference position and, then, the stimulus was presented. The listeners were asked to point to the perceived stimulus direction. Each condition was tested in three experimental blocks each. An experimental block consisted of 100 targets and lasted for approximately 20 min. The order of the 21 experimental blocks was randomized in such a way that within seven blocks all conditions were in a randomized order. The study was a double-blind experiment, i.e., none of the listeners were the authors, listeners were not enlightened as to the nature of the experiment, listeners and the experimenter did not know the tested conditions. As in Sec. IV, the PE and the QE were calculated for the analysis of localization performance. Additionally, the lateral RMS error (LE) and the signed polar bias were calculated.

B. Results and discussion

1. General

Figure 11 shows the localization performance parameters PE and QE obtained from the localization experiment (labeled as Original). Generally, an increase of the AEL yielded degraded performance (1–5 in Fig. 11), more or less approaching that obtained for the KEMAR HRTFs (K in Fig. 11).

For the numerically calculated HRTFs with AELs of 1 and 2 mm, both PE and QE were in the range usually found

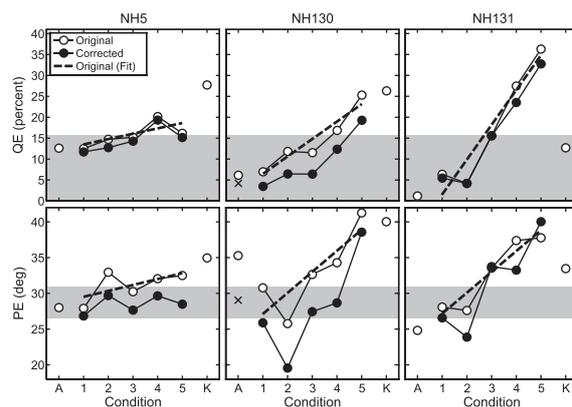


FIG. 11. Localization performance observed for NH5, NH130, and NH131 in the localization experiment (Original). Conditions: acoustically measured own-ear DTFs (labeled as A), numerically calculated own-ear DTFs for the coarsened meshes (1–5), and acoustically measured manikin DTFs of the KEMAR (K). The crosses represent results from the acoustic training for NH130. Original (Fit): linear fit to the results. Corrected: Original results recalculated with corrected polar bias. Grey area as in Fig. 10.

in localization experiment with listener-specific HRTFs (e.g., Middlebrooks, 1999a). These two conditions yielded also QEs very similar to those obtained in the condition with acoustically measured HRTFs (A in Fig. 11), for all three listeners. For the PEs, the situation was different: the PEs were similar for NH5 and NH131, but not for NH130.

NH130 showed actually a surprisingly large PE in the condition with acoustically measured HRTFs. Since the PEs in the conditions with HRTFs numerically calculated for AELs of 1 and 2 mm were within the expected range, we looked deeper into this issue. We identified four issues which might have yielded the poor PE in the condition with acoustically measured HRTFs. First, NH130 might generally have been a poor localizer. In that case, however, NH130 would not be able to perform such good in the conditions with numerically calculated HRTFs. Second, the acoustically measured HRTFs might have been problematic. We used exactly the same HRTFs in both training and main experiment, and the PE and QE from the training was 29.1° and 4.7%, respectively, showing that NH130 did much better in the training than in the later experiment. The performance from the training was in the range of usually observed localization performance, indicating that, in general, NH130 was able to localize sounds quite well even with the acoustically measured HRTFs. Third, in the main experiment, seven conditions were tested, potentially offering room for order effects like practice or fatigue. However, in the experiment, each of the seven conditions was split in three blocks, the blocks were presented in a randomized order, and the obtained PE was consistently poor in all three corresponding blocks. Thus, the order effect was unlikely the reason for poor performance with acoustically measured HRTFs. Fourth, we noticed that NH130 showed a large variability in PE across all tested conditions. In default of a better hypothesis, we can only speculate: While NH130, being a general good localizer, seems to have produced an outlier in that one particular condition. Anyway, from now on, we rely more on

NH130's results from the acoustic training, which seem to be more reasonable (see the crosses in Fig. 11).

The performance of NH5 did not change much as a function of AEL. NH5 seems thus to be not that susceptible to AEL increase as NH130 and NH131 seem to be. However, the localization error obtained for NH5 in the acoustic and 1-mm conditions was larger than compared to the other subjects (with the PE of NH130 from the training as discussed before). This renders NH5 a generally less sensitive localizer. There is evidence that listener-specific sound-localization performance is linked to the ability to discriminate spectral shapes (Andéol *et al.*, 2013). Such a factor seems to be even more salient than the directional features provided by listener-specific HRTFs (Majdak *et al.*, 2014). Thus, for an insensitive subject as NH5 might be, the degradation of spectral features resulting from the AEL increase might have little impact on the sound-localization performance. On the other hand, NH131 seems to be a sensitive subject because a small increase of AEL yielded large degradation of the localization performance. The sensitivity of NH130 seems to be between that of NH131 and NH5. As a common ground, HRTFs calculated with AELs of 1 or 2 mm seem to yield good localization performance in all our subjects.

In order to quantify the effect of AEL on the localization performance, a linear regression was fit to PEs and QEs obtained for AELs of 1–5 mm. For all subjects, the linear regression showed a positive slope, i.e., a degrading localization performance for increasing AEL in the mesh. The PE increased by 0.8°, 2.9°, and 2.9°/mm for NH5, NH130, and NH131, respectively, and 2.2°/mm on average. The QE increased by 1.3%, 4.2%, and 8.3%/mm for NH5, NH130, and NH131, respectively, and 4.6%/mm on average.

2. Polar bias and the interaural horizontal plane

The polar bias in localization responses was calculated for the conditions involving numerically calculated HRTFs. For NH5, NH130, and NH131, the bias was -8.1° , -20.1° , and -11.6° , respectively, and -13.3° on average. This finding implies that our definition of the interaural horizontal plane as the intersection of the nasal and auricular points (see Sec. IID) did not correspond to the interaural horizontal plane used by the listeners in the experiment. This finding is actually not surprising because distinct markers of the interaural horizontal plane for a good match between the acoustical measurement and mesh generation were not available. For example, Jin *et al.* (2014) solved this problem *post hoc* by applying a rotation minimizing the spectral cross-correlation between the acoustically measured and numerically calculated HRTFs. The average rotation they had applied was 12.6° .

Given these findings, we were looking for a good definition of the interaural horizontal plane in the process of mesh generation. In the field of anatomy, besides the nasal points, the orbital points, i.e., the inferior margins of the orbits, can be used for alignment of the human head. Further, the plane intersecting orbital and auricular points, called the Frankfurt plane, is used to describe the position of head usually carried

in the living subjects (Cheng *et al.*, 2012). The Frankfurt plane, together with the median plane can be used to define the three anatomical orientations: lateral, anterior–posterior (front–back), and superior–inferior (up–down). Thus, we checked the position of the Frankfurt plane in our meshes.

For NH5, NH130, and NH131, the Frankfurt plane deviated from our previous definition of the interaural horizontal plane by 10.5° , 15.9° , and 13.4° , respectively, with an average of 13.3° . This deviation might explain the polar bias in the localization responses.

In order to apply the Frankfurt plane as horizontal plane to our localization responses, for each sound target, we subtracted the deviation to the Frankfurt plane from the target angle. This polar shift corresponds to a rotation of the mesh in the numerical calculation similar to that from Jin *et al.* (2014). For NH5, NH130, and NH131, the resulting corrected polar bias was 2.4° , -4.2° , and 1.8° , respectively, and 0.0° on average.

We then recalculated the localization performance given the corrected target angles (see Fig. 11, data labeled as Corrected). The smaller corrected PEs and QEs indicate that the original PEs and QEs were indeed subject of the polar bias resulting from the non-optimal definition of the interaural horizontal plane used in the first place.

In summary, both the vanishing polar bias and the smaller corrected PEs and QEs suggest that the Frankfurt plane is appropriate for the definition of the interaural horizontal plane in the process of mesh alignment for numerical HRTF calculation.

3. Lateral errors

Table V shows the lateral RMS error (LE) for all three subjects and statistics of the LE observed in virtual sound-localization experiments from Middlebrooks (1999a). For the 1-mm condition, the LE was in the range of LEs from Middlebrooks (1999a). It seems that the AEL had a negligible effect on LE, confirming predictions from the TOA model. The LE for NH131 was larger for numerically calculated HRTFs than for acoustically measured HRTFs but still was in the range of parameters found by (Middlebrooks, 1999a). Since impedance boundary conditions can affect the ITD (Katz, 2001b), this difference might be explained by the discrepancy of actual boundary conditions of a subject's

TABLE V. Lateral RMS error (LE). Middlebrooks: Average results from Middlebrooks (1999a).

	LE (°)		
	NH5	NH130	NH131
AC	15.8	17.6	9.3
1 mm	13.2	15.3	13.1
2 mm	16.4	15.5	12.1
3 mm	16.1	19.0	14.8
4 mm	15.5	17.3	15.6
5 mm	16.2	16.8	14.9
Middlebrooks		14.5 ± 2.2	
KEMAR	17.5	17.7	11.6

head and boundary conditions used in the numerical calculation.

4. Predictions and actual performance

Looking back to the predictions from Sec. IV, similar degradation of localization performance with increasing AEL was observed in the predicted and the actual localization performance. The Pearson's correlation coefficients between predicted and actual localization-performance parameters were 0.73 and 0.83 for PE and QE, respectively.

Even though there seems to be a bias between the predictions and the actual performance, the trend in the performance degradation found in actual experiments seems to be reflected by the model predictions; an aspect important for the correct interpretation of predictions from the previous sections. If required in the future, with the actual responses from the localization experiment, the model can be calibrated to the actual listener-specific uncertainties, further improving the congruency between our predictions and the actual performance (Majdak *et al.*, 2014).

VI. CONCLUSIONS

We investigated requirements on numerical calculation of listener-specific HRTFs for accurate sound-localization performance. We focused on the size and position of the virtual microphone and on the discretization of the listener's head and pinna geometry.

A simple model of the microphone used in acoustic measurements seems to be sufficient. A single mesh element, approximately centered at the blocked ear canal yielded HRTFs comparable to those from acoustic measurements and more sophisticated virtual microphones were not required. With a tolerance of $PE \pm 0.5^\circ$ and $QE \pm 0.5\%$, any position of the virtual microphone within the distance of ± 2.5 mm around the ear-canal center seems to be acceptable. Larger virtual microphones were more susceptible to their exact position.

The resolution of the discretized listener's geometry had an obvious impact on the sound-localization performance in sagittal planes (top/down, front/back). The degradation of the geometrical detail level (geometrical error) was the dominant factor. The under-sampling of the sound-field (numerical error) did not affect our results much. The localization performance degraded with increasing AEL, with an average rate of $2.2^\circ/\text{mm}$ and $4.6\%/\text{mm}$ (PE per AEL and QE per AEL, respectively). For AELs of 1 and 2 mm, the localization performance was in the range of that usually found when localizing virtual sound sources.

The resolution of the discretized geometry had a negligible effect on the sound-localization performance in lateral dimension (left/right). This result is plausible, because the change of the AEL was small compared to the dimension of the human head, thus, having a negligible effect on shadowing and diffraction caused by the head.

The analysis of the polar bias indicates that the perceived interaural horizontal plane (0° elevation) in the mesh generation process is better represented by the Frankfurt plane than by the tip of the nose.

Some of our findings are based on model predictions only. For conditions which were investigated by both model predictions and actual sound-localization experiments, we found that the predicted localization performance corresponded well (average correlation coefficient of 0.78) to the performance obtained in the actual localization experiments. This seems to be a good basis for drawing conclusions based on the predictions only.

In summary, we hope that our findings help in further establishing methods for the numerical HRTF calculation. While our study is clearly limited to sound localization and to three subjects only, further investigation of the requirements for other perceptual effects like sound externalization and timbre seem to be the next logical step.

ACKNOWLEDGMENTS

Special thanks go to Andreas Reichinger (VRVis GmbH, Austria) and his team for the work on geometry acquisition and generation of the reference mesh. We are also grateful to our subjects for their patience during the measurements and experiments. This study was supported by the Austrian Science Fund (FWF, P 24124-N13).

¹Some spectral features below 5 kHz are caused by reflections at the torso (Algazi *et al.*, 2001b), however, these low frequency components do not contribute much to the localization of broadband signals (Bronkhorst, 1995).

²HRTFs and anthropometric data available from <http://www.kfs.oew.ac.at/hrtf/> (Last viewed June 24, 2014).

³Description available from http://www.kfs.oew.ac.at/research/experimental_audiology/hrtf/readme.pdf (Last viewed June 24, 2014).

⁴The software package Mesh2HRTF was used, available from <http://mesh2hrtf.sourceforge.net/> (Last viewed June 12, 2015).

⁵Specifications available from http://www.zcorp.com/documents/491_8749-ZScanner700_CX_SpecSheet.pdf (Last viewed June 24, 2014).

⁶Specifications available from <http://www.ge-mcs.com/en/radiography-xray/ct-computed-tomography/v-tome-x-c.html> (Last viewed June 24, 2014).

⁷Geomagic, available from <http://www.geomagic.com/> (Last viewed June 24, 2014).

⁸Since a subject's head was not fixed for the acoustic measurement, the head alignment to the loudspeaker grid was not accurately controllable. Subjects were instructed to sit comfortable and to look at the loudspeaker located at 0° azimuth and 0° elevation. A subject's head was tracked during the measurement and a rotation of $\pm 2^\circ$ was allowed.

⁹Meshlab, available from <http://meshlab.sourceforge.net> (Last viewed June 24, 2014).

¹⁰OpenFlipper, available from <http://www.openflipper.org> (Last viewed July 8, 2014).

¹¹Retrieved from <http://sofacoustics.org/data/database/ari/> (Last viewed January 20, 2015), originating from the "(short)" HRTF set from <http://www.kfs.oew.ac.at/hrtf> (Last viewed January 20, 2015).

¹²Retrieved from <http://sofacoustics.org/data/database/cipic/> (Last viewed January 20, 2015), originating from <http://interface.cipic.ucdavis.edu/sound/hrtf.html> (Last viewed January 20, 2015).

¹³Retrieved from <http://sofacoustics.org/data/database/listen/> (Last viewed January 20, 2015), originating from the "compensated" HRTF set from <http://recherche.ircam.fr/equipements/salles/listen/download.html> (Last viewed January 20, 2015).

¹⁴Note that the minimum-phase cross-correlation method showed the best correspondence to actual radii in numerically calculated HRTFs (Ziegelwanger and Majdak, 2014), the results for other methods may vary (Katz and Noisternig, 2014).

Algazi, V. R., Avendano, C., and Duda, R. O. (2001a). "Estimation of a spherical-head model from anthropometry," *J. Audio Eng. Soc.* **49**, 472–479.

- Algazi, V. R., Avendano, C., and Duda, R. O. (2001b). "Elevation localization and head-related transfer function analysis at low frequencies." *J. Acoust. Soc. Am.* **109**, 1110–1122.
- Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (2001c). "The CIPIC HRTF database," in *Proceedings of 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, New York, pp. 99–102.
- Andéol, G., Macpherson, E. A., and Sabin, A. T. (2013). "Sound localization in noise and sensitivity to spectral shape," *Hear. Res.* **304**, 20–27.
- Andreopoulou, A., Begault, D., and Katz, B. (2015). "Inter-laboratory round robin HRTF measurement comparison," *IEEE J. Sel. Top. Signal Process.* **PP**(99), 1–12.
- Baumgartner, R., Majdak, P., and Laback, B. (2013). "Assessment of sagittal-plane sound localization performance in spatial-audio applications," in *The Technology of Binaural Listening*, edited by J. Blauert (Springer, Berlin), Chap. 4.
- Botsch, M., and Kobbelt, L. (2004). "A remeshing approach to multiresolution modeling," in *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing (SGP'04)* (ACM, New York), pp. 185–192.
- Bronkhorst, A. W. (1995). "Localization of real and virtual sound sources," *J. Acoust. Soc. Am.* **98**, 2542–2553.
- Chen, Z.-S., Waubke, H., and Kreuzer, W. (2008). "A formulation of the fast multipole boundary element method (FMBEM) for acoustic radiation and scattering from three-dimensional structures," *J. Comput. Acoust.* **16**, 303–320.
- Cheng, Y., Leow, W.-K., and Lim, T. C. (2012). "Automatic identification of Frankfurt plane and mid-sagittal plane of skull," in *Proceedings of 2012 IEEE Workshop on the Applications of Computer Vision*, Breckenridge, CO, pp. 233–238.
- Cignoni, P., Rocchini, C., and Scopigno, R. (1998). "Metro: Measuring error on simplified surfaces," *Comput. Graph. Forum* **17**, 167–174.
- Greff, R., and Katz, B. F. G. (2007). "Round robin comparison of HRTF simulation systems: Preliminary results," in *Proceedings of the 123rd Convention of the Audio Engineering Society*, New York, Convention Paper 7188.
- Gumerov, N. A., O'Donovan, A. E., Duraiswami, R., and Zotkin, D. N. (2010). "Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation," *J. Acoust. Soc. Am.* **127**, 370–386.
- Hebrank, J., and Wright, D. (2005). "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Am.* **56**, 1829–1834.
- Huttunen, T., Seppälä, E. T., Kirkeby, O., Kärkkäinen, A., and Kärkkäinen, L. (2007). "Simulation of the transfer function for a head-and-torso model over the entire audible frequency range," *J. Comput. Acoust.* **15**, 429–448.
- Jin, C. T., Guillon, P., Epain, N., Zolfaghari, R., van Schaik, A., Tew, A. I., Hetherington, C., and Thorpe, J. (2014). "Creating the Sydney York morphological and acoustic recordings of ears database," *IEEE Trans. Multimedia* **16**, 37–46.
- Kahana, Y., and Nelson, P. A. (2006). "Numerical modelling of the spatial acoustic response of the human pinna," *J. Sound Vib.* **292**, 148–178.
- Katz, B. F. G. (2001a). "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *J. Acoust. Soc. Am.* **110**, 2440–2448.
- Katz, B. F. G. (2001b). "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," *J. Acoust. Soc. Am.* **110**, 2449–2455.
- Katz, B. F. G., and Noisternig, M. (2014). "A comparative study of interaural time delay estimation methods," *J. Acoust. Soc. Am.* **135**, 3530–3540.
- Kreuzer, W., Majdak, P., and Chen, Z. (2009). "Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range," *J. Acoust. Soc. Am.* **126**, 1280–1290.
- Kuhn, G. F. (1977). "Model for the interaural time differences in the azimuthal plane," *J. Acoust. Soc. Am.* **62**, 157–167.
- Langendijk, E. H. A., and Bronkhorst, A. W. (2002). "Contribution of spectral cues to human sound localization," *J. Acoust. Soc. Am.* **112**, 1583–1596.
- Lopez-Poveda, E. A., and Meddis, R. (1996). "A physical model of sound diffraction and reflections in the human concha," *J. Acoust. Soc. Am.* **100**, 3248–3259.
- Lorensen, W. E., and Cline, H. E. (1987). "Marching cubes: A high resolution 3D surface construction algorithm," in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'87)*, edited by Maureen C. Stone (ACM, New York), pp. 163–169.
- Macpherson, E. A., and Middlebrooks, J. C. (2002). "Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited," *J. Acoust. Soc. Am.* **111**, 2219–2236.
- Majdak, P., Baumgartner, R., and Laback, B. (2014). "Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization," *Front. Psychol.* **5**(319), 1–10.
- Majdak, P., Goupell, M. J., and Laback, B. (2010). "3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Atten. Percept. Psychophys.* **72**, 454–469.
- Majdak, P., Masiero, B., and Fels, J. (2013). "Sound localization in individualized and non-individualized crosstalk cancellation systems," *J. Acoust. Soc. Am.* **133**, 2055–2068.
- Marburg, S. (2002). "Six boundary elements per wavelength: Is that enough?," *J. Comput. Acoust.* **10**, 25–51.
- Mey, F. D., Reijnen, J., Peremans, H., Otani, M., and Firzloff, U. (2008). "Simulated head related transfer function of the phyllostomid bat *Phyllostomus discolor*," *J. Acoust. Soc. Am.* **124**, 2123–2132.
- Middlebrooks, J. C. (1999a). "Virtual localization improved by scaling non-individualized external-ear transfer functions in frequency," *J. Acoust. Soc. Am.* **106**, 1493–1510.
- Middlebrooks, J. C. (1999b). "Individual differences in external-ear transfer functions reduced by scaling in frequency," *J. Acoust. Soc. Am.* **106**, 1480–1492.
- Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (1995). "Head-related transfer functions of human subjects," *J. Audio Eng. Soc.* **43**, 300–321.
- Morse, P. M., and Ingard, K. U. (1986). *Theoretical Acoustics* (Princeton University Press, Princeton, NJ), pp. 1–949.
- Rébillat, M., Benichoux, V., Otani, M., Keriven, R., and Brette, R. (2014). "Estimation of the low-frequency components of the head-related transfer functions of animals from photographs," *J. Acoust. Soc. Am.* **135**, 2534–2544.
- Reichinger, A., Majdak, P., Sablatnig, R., and Maierhofer, S. (2013). "Evaluation of Methods for Optical 3-D Scanning of Human Pinna," in *Proceedings of the International Conference on 3D Vision (3DV 2013)*, pp. 390–397.
- Sauter, S. A., and Schwab, C. (2010). *Boundary Element Methods* (Springer, Heidelberg), pp. 1–500.
- Takemoto, H., Mokhtari, P., Kato, H., Nishimura, R., and Iida, K. (2012). "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," *J. Acoust. Soc. Am.* **132**, 3832–3841.
- Treeby, B. E., and Pan, J. (2009). "A practical examination of the errors arising in the direct collocation boundary element method for acoustic scattering," *Eng. Anal. Boundary Elem.* **33**, 1302–1315.
- Treeby, B. E., Pan, J., and Paubally, R. M. (2007a). "An experimental study of the acoustic impedance characteristics of human hair," *J. Acoust. Soc. Am.* **122**, 2107–2117.
- Treeby, B. E., Pan, J., and Paubally, R. M. (2007b). "The effect of hair on auditory localization cues," *J. Acoust. Soc. Am.* **122**, 3586–3597.
- Wightman, F. L., and Kistler, D. J. (1989). "Headphone simulation of free-field listening. I: Stimulus synthesis," *J. Acoust. Soc. Am.* **85**, 858–867.
- Ziegelwanger, H., and Majdak, P. (2014). "Modeling the direction-continuous time-of-arrival in head-related transfer functions," *J. Acoust. Soc. Am.* **135**, 1278–1293.
- Ziegelwanger, H., Reichinger, A., and Majdak, P. (2013). "Calculation of listener-specific head-related transfer functions: Effect of mesh quality," *POMA* **19**, 050017.

Chapter 6

Non-uniform sampling of geometry for the numeric simulation of HRTFs

This work was published as:

Ziegelwanger, H., Majdak, P., and Kreuzer, W. (2014). "Non-uniform sampling of geometry for the numeric simulation of head-related transfer functions," in *Proceedings of the 21st International Congress of Sound and Vibration*, Beijing, CN, 1-8. [doi:10.13140/RG.2.1.4590.6968](https://doi.org/10.13140/RG.2.1.4590.6968)

The non-uniform re-meshing of geometry for the numerical HRTF calculation was my idea. In collaboration with the co-authors, an argumentation line to justify the method was designed. I, as the first author, numerically calculated and acoustically measured the HRTFs, analyzed the HRTFs in terms of model predictions, generated all figures and wrote the initial draft of the manuscript. The co-authors helped writing the manuscript.

Co-author 1:

Co-author 2:



NON-UNIFORM SAMPLING OF GEOMETRY FOR THE NUMERIC SIMULATION OF HEAD-RELATED TRANSFER FUNCTIONS

Ziegelwanger, Harald; Majdak, Piotr; Kreuzer, Wolfgang

Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria 1040

e-mail: harald.ziegelwanger@oeaw.ac.at

Head-related transfer functions (HRTFs) describe the directional filtering effect caused by the morphology of a listener's head, torso, and pinna. While HRTFs are usually measured in an acoustic procedure, they can also be numerically simulated with the fast-multipole boundary-element method. To this end, the exact geometrical data of a listener's head and pinna are required to obtain perceptively valid HRTFs and are usually acquired in visual scans. The scanned surface geometry is then discretized yielding a three-dimensional polygon mesh. In general, for the discretization, at least six elements per wavelength are recommended in order to keep the numerical error small and the calculation stable. When applying this general recommendation to HRTFs, the corresponding mesh results in a densely populated linear equation system with up to 150 000 degrees of freedom and the simulation process lasts tens of hours. In this study, we show that due to perceive direction-dependent requirements on the accuracy of HRTFs, the recommended number of elements can be substantially reduced in some areas of the mesh, resulting in a non-uniform mesh discretization. We further show the discretization conditions under which a significant increase in computational efficiency can be achieved without introducing a noticeable change in the sound localization performance when listening with the simulated HRTFs.

1. Introduction

Head related transfer functions (HRTFs) describe the directional filtering of the incoming sound at a listener's ear canal¹ and they are essential for binaural audio reproduction systems². HRTFs are usually acquired in an acoustic measurement procedure^{3,4}, in which small microphones are placed into listener's ear canals and transfer functions are measured in an anechoic chamber for many directions around the listener.

Numerical simulation of the HRTFs has received much attention in recent years⁵⁻¹⁰. In contrast to the acoustic measurement, the numerical simulation requires a digital representation of a listener's geometry only, i.e., a three-dimensional surface mesh. The simulation is usually done by using the boundary-element method (BEM), in which at least six sampling points per wavelength uniformly distributed over the surface geometry are recommended to retrieve accurate results¹¹. Because the matrix describing the linear system of equations generated by the BEM is dense, the computational costs increase at least quadratically with the number of sampling points. The computation time can be reduced by modeling the in-ear microphone as sound source and evaluating the sound field at the loudspeaker positions, i.e. the evaluation nodes, which is equivalent to the reciprocal acoustic procedure¹². Here, the ipsilateral ear is the ear where the sound source is placed and the

contra-lateral ear is the ear at the opposite side of the head. In order to further reduce the computation time and the amount of required memory, the BEM can be coupled with the fast-multipole method (FMM)^{9,10}. But, still, there are 150 000 unknowns and HRTF simulation for the whole audible frequency range lasts tens of hours on a single CPU¹⁰.

The computational costs and the number of unknowns can be dramatically reduced by representing the geometry with less sampling points at carefully chosen positions. This, however, violates the generally applied recommendation to use at least six elements per wavelength.

In this study, we justify such a violation at special regions of the surface and consider a *non-uniform* representation of the underlying geometry allowing for a smaller matrix in the simulations. Further, HRTFs were simulated using a *non-uniformly* sampled geometry of the head of a human listener. The resulting HRTFs were compared to acoustically measured HRTFs and to HRTFs simulated for a uniformly sampled geometry of the same listener. The quality of the HRTFs was evaluated in terms of the modeled sound-localization performance in sagittal planes¹³.

2. Boundary Element Method

The main equation for acoustics in an exterior domain Ω_e outside an object Ω with boundary Γ and Neumann boundary conditions is the Helmholtz equation:

$$\begin{aligned} \nabla^2 \phi(\mathbf{x}) + k^2 \phi(\mathbf{x}) &= 0, \quad \mathbf{x} \in \Omega_e \\ \lim_{|\mathbf{x}| \rightarrow \infty} |\mathbf{x}| \left(\frac{\partial \phi(\mathbf{x})}{\partial |\mathbf{x}|} - ik \phi(\mathbf{x}) \right) &= 0 \\ \frac{\partial \phi(\mathbf{x})}{\partial \mathbf{n}_x} &= v(\mathbf{x}), \quad \mathbf{x} \in \Gamma, \end{aligned} \quad (1)$$

where $\phi(\mathbf{x}) = \frac{p}{i\omega\rho}$ denotes the velocity potential at the point \mathbf{x} and $k = \frac{\omega}{c}$ is the wavenumber dependent on the speed of sound c and the circular frequency ω . p represents the sound pressure, ρ the density of the medium and i the imaginary unit. Γ is the boundary surface (here, head and ears), \mathbf{n} the vector normal to Γ at a point \mathbf{x} and v the particle velocity in direction of \mathbf{n} . To avoid problems at irregular frequencies the approach of Burton and Miller¹⁴ was used. With the acoustic principle of reciprocity the final integral equation is:

$$-\frac{1}{2} \phi(\mathbf{x}) + L[\phi](\mathbf{x}) + \frac{i}{k} \frac{\partial}{\partial \mathbf{n}_x} L[\phi](\mathbf{x}) = \frac{i}{k} \frac{1}{2} v(\mathbf{x}) + \tilde{L}[v](\mathbf{x}) + \frac{i}{k} \frac{\partial}{\partial \mathbf{n}_x} \tilde{L}[v](\mathbf{x}), \quad \mathbf{x} \in \Gamma, \quad (2)$$

with

$$\begin{aligned} L[\phi](\mathbf{x}) &:= \int_{\Gamma} H(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) d\mathbf{y} \\ \tilde{L}[v](\mathbf{x}) &:= \int_{\Gamma} G(\mathbf{x}, \mathbf{y}) v(\mathbf{y}) d\mathbf{y}, \end{aligned} \quad (3)$$

and $G(\mathbf{x}, \mathbf{y})$ and $H(\mathbf{x}, \mathbf{y})$ as the Green's function of the Helmholtz equation and its derivative with respect to the normal vector, respectively:

$$\begin{aligned} G(\mathbf{x}, \mathbf{y}) &:= \frac{e^{ikr}}{4\pi r}, \\ H(\mathbf{x}, \mathbf{y}) &:= \frac{\partial G(\mathbf{x}, \mathbf{y})}{\partial \mathbf{n}_y} = \frac{e^{ikr}}{4\pi r} \left(ik - \frac{1}{r} \right) \frac{\partial r}{\partial \mathbf{n}_y}, \\ r &:= \|\mathbf{y} - \mathbf{x}\|. \end{aligned} \quad (4)$$

Figure 1 shows the amplitude level of Green's function, G , and its derivative, H , as functions of the distance, r , for various frequencies. Note the singularity at $r=0$ and the fast decay.

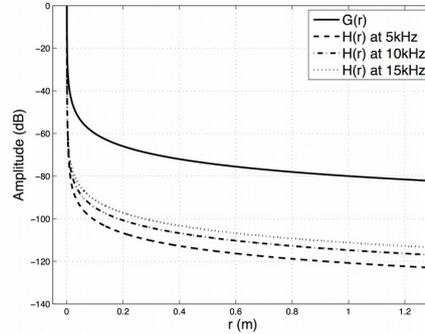


Figure 1. Green's function and its derivative as function of the distance r . Normalized at $r = 1$ mm.

Sampling of the surface Γ yields a mesh consisting of a set of triangular elements Γ_j . By using collocation with constant boundary elements¹⁵, Eq. 2 converts into a sum of integrals which can be written as a system of equations:

$$\mathbf{A}\boldsymbol{\phi} = \mathbf{b}, \quad (5)$$

where $\boldsymbol{\phi}$ is the vector of unknown velocity potentials ϕ_j at the midpoints of the elements, i.e. the collocation nodes,

$$A_{i,j} = -\frac{1}{2}\delta_{i,j} + \int_{\Gamma_j} H(\mathbf{x}_i, \mathbf{y}) d\mathbf{y} + \frac{i}{k} \int_{\Gamma_j} \frac{\partial}{\partial \mathbf{n}_x} H(\mathbf{x}_i, \mathbf{y}) d\mathbf{y}, \quad (6)$$

and

$$b_i = \frac{i}{2k} v_i + \sum_j v_j \int_{\Gamma_j} G(\mathbf{x}_i, \mathbf{y}) d\mathbf{y} + \frac{i}{k} \sum_j v_j \int_{\Gamma_j} \frac{\partial}{\partial \mathbf{n}_y} G(\mathbf{x}_i, \mathbf{y}) d\mathbf{y}, \quad (7)$$

δ denotes the Kronecker delta function, \mathbf{x}_i are the collocation nodes, v_i and v_j are the constant values of the particle velocity at element Γ_i and Γ_j , respectively.

The solution for the vector $\boldsymbol{\phi}$ is calculated by using an iterative solver. Regular integrals are solved using standard Gauss-Legendre quadrature, for weakly singular integrals, i.e. when the kernel functions become singular, the method of singular local coordinate transformation is applied. For a detailed description refer to Ref. 16 and 17. To reduce the computational costs in the simulation process, the BEM was coupled with the FMM. For further details refer to Ref. 10.

With $\boldsymbol{\phi}$, any point \mathbf{x} in the exterior domain Ω_e can be evaluated:

$$\phi(\mathbf{x}) = \sum_j \left[\int_{\Gamma_j} H(\mathbf{x}, \mathbf{y}) d\mathbf{y} \phi_j - \int_{\Gamma_j} G(\mathbf{x}, \mathbf{y}) d\mathbf{y} v_j \right]. \quad (8)$$

3. Geometry

3.1 Visual Acquisition

The head of a male human subject was scanned by a laser scanner (ZScanner 700CXi, ZCorp). The mesh was extracted from an internal volumetric representation with approximately 0.5 mm voxel size. The neck was cut, rounded, and closed.

In order to obtain a very detailed representation of the pinna, a high-energy industrial computer-tomography scanner was used to capture the geometry of silicone impressions of the ears. The impressions were scanned using 1000 X-ray images spaced in 0.36° intervals. The iso-surfaces separating air and silicone were extracted using auto-thresholding, i.e., a proprietary version of the marching cubes algorithm¹⁸.

The pinna meshes were merged with the head mesh by manually aligning them with the pinnae of the head mesh via tie points (Manual Registration, Geomagic) and refining (Best Fit Alignment, Geomagic). The seam-region was re-sculpted and smoothed (Voxel-Sculpting, 3D-Coati) in a resolution of below 0.5 mm voxel-size. The final high-accuracy mesh was water-tightened and consisted of approximately 2.5 million triangles. For further details please refer to Ref. 19.

3.2 Uniformly Sampled Geometry

Because of the memory limits of our computers, the number of elements used in our computer model was limited to approximately 120 000 elements. Thus, the high-accuracy mesh was approximately uniformly resampled to an average edge length of 2 mm (Openflipper²⁰), which follows the recommendation of 6 elements per wavelength for frequencies up to 20 kHz (Figure 2 b).

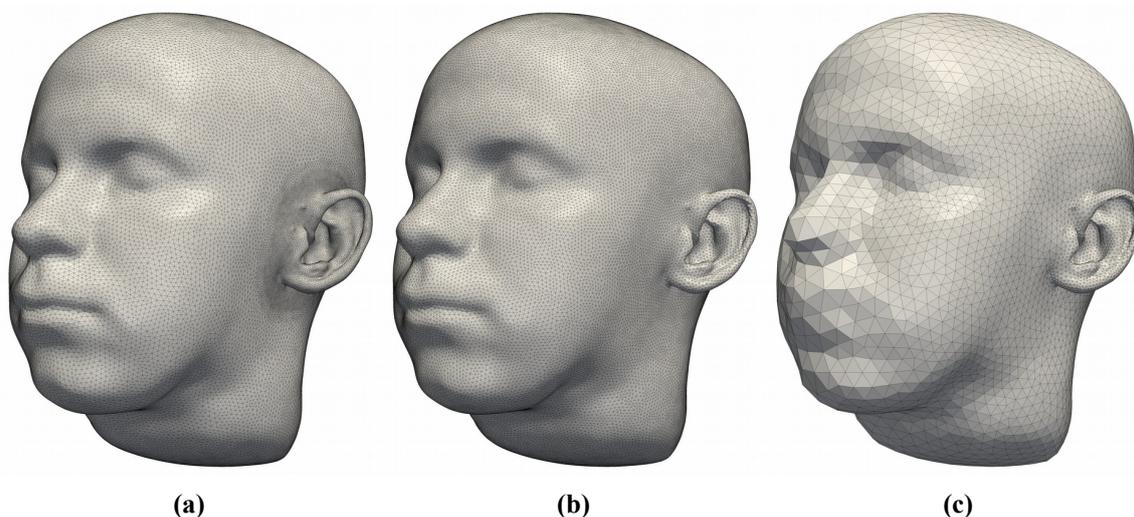


Figure 2. Renderings of the reference object (a), the uniformly sampled geometry with an average edge length of 2 mm (b), and the non-uniformly sampled geometry with an edge length ranging from 1 to 25 mm (c).

3.3 Non-uniformly Sampled Geometry

We aim at reducing the number of elements in certain regions of the mesh. We assume, that increasing the element size for regions far apart from the modeled sound source in the ipsilateral ear canal does not effect the resulting HRTFs. We motivate the loosening of the recommendation of at least 6 elements per wavelength as follows.

First, the Green's function and its derivatives decay fast with increasing distance. Thus, errors arising from evaluating Eq. 6 for boundary elements far apart from the modeled sound source barely affect the results of the evaluation of Eq. 6 for boundary elements near the modeled sound source.

Second, the distance between the boundary elements and the evaluation nodes is usually in the order of 1 m. At that distance, the magnitudes of G and H change less than 5 dB when evaluated in the proximity of around 10 cm (i.e., the typical head radius). Thus, an error due to the increased element size will affect the integral in Eq. 8 for all HRTF-directions at a similar level. However, the surface sound pressure decreases down to -90 dB at the contralateral ear compared to that at the sound source. Thus, the error due to the increased element size at the contralateral side will be damped in the order of tens of dB, having only a small effect on the final results (HRTFs).

Non-uniformly sampled meshes of the head were created based on the high-accuracy mesh. First, the contralateral ear was cut out, the resulting hole was filled and the surface of the head excluding the ipsilateral pinna was smoothed (Sculptris). Note, that two meshes are required to calcu-

late HRTFs for both ears. Then, the surface was non-uniformly resampled based on a curvature based re-meshing algorithm (Openflipper²⁰) for different parameter sets. The minimal edge length at the pinna was fixed at 1 mm and the maximal edge length at the contralateral head side was varied from 4 to 40 mm. Geometry parameters for an exemplary non-uniformly sampled geometry are listed in Table 1. The corresponding head mesh is shown in Figure 2 (c) and the pinna is shown in Figure 3 (c).

4. Results

4.1 Acoustic Measurement

Acoustic measurement was done in a semi-anechoic chamber³. HRTFs for 1550 directions were measured by placing in-ear-microphones in the subject's ear canals, i.e., blocked ear-canal method⁴. The total measurement time was reduced to 20 minutes by the multiple exponential-sweep method²¹. The horizontal and vertical range was 360° and -30° to 80°, respectively. The horizontal resolution was 2.5° and 5° for directions inside and outside of ±45°, respectively. The vertical resolution was 5°. The measured transfer functions were equalized with the equipment transfer functions by spectral division. Then, directional transfer functions (DTFs) were calculated²², which were temporally windowed with an asymmetric Tukey window to a length of 256 samples (5.33 ms at a sampling rate of 48 kHz).

Figure 3 (d) shows the amplitude spectra of DTFs in the median-plane. The black color represents notches, which are assumed to be relevant for the sound-source localization in sagittal planes. For that DTFs, the sound localization performance was modeled with a sound-localization model for sagittal planes¹³. In that model, the localization performance is represented by the amount of front-back confusions (quadrant error) and the local localization precision (polar error), which was for our DTFs 6.4 % and 27°, respectively. Figure 3 (g) shows the energy time curves of HRTFs in the horizontal plane. The black color represents the energy of the impulse response. The onset of the impulse responses represent the time-of-arrivals. The interaural difference of time-of-arrivals is the interaural time difference, which is relevant for the lateralization of a sound-source²³.

Table 1. Mesh properties and localization performance

	Measured	Uniform	Non-uniform
Number of Elements		98 564	8 968
Min. Edge Length (mm)		2	1
Max. Edge Length (mm)		2	25
Avg. Edge Length (mm)		2	5
Computation Time (h)		14.96	1.05
Polar RMS error (deg)	27	28.9	28.9
Quadrant error rate (%)	6.4	10.3	11.3

4.2 Numeric Calculations

The head was modeled rigid and the microphone was modeled as sound source by applying a velocity boundary condition:

$$v(\mathbf{x}) = v_0(\mathbf{x}) = \begin{cases} v_{mic} & \mathbf{x} \in \Gamma_M \\ 0 & \text{otherwise} \end{cases}, \quad (9)$$

where Γ_M denotes the area of the in-ear microphone placed in the ear canal and v_{mic} was set to 0.1 m/s. Computation was done on a Linux cluster containing eight machines with Intel i7-3820 processors running at 3.6 GHz and 64 GB of RAM each. The complex sound pressure on the boundary surface was calculated at 200 sampling points in the frequency domain, which were linearly spaced between 0.1 and 20 kHz. Calculation time for the uniformly sampled geometry and an

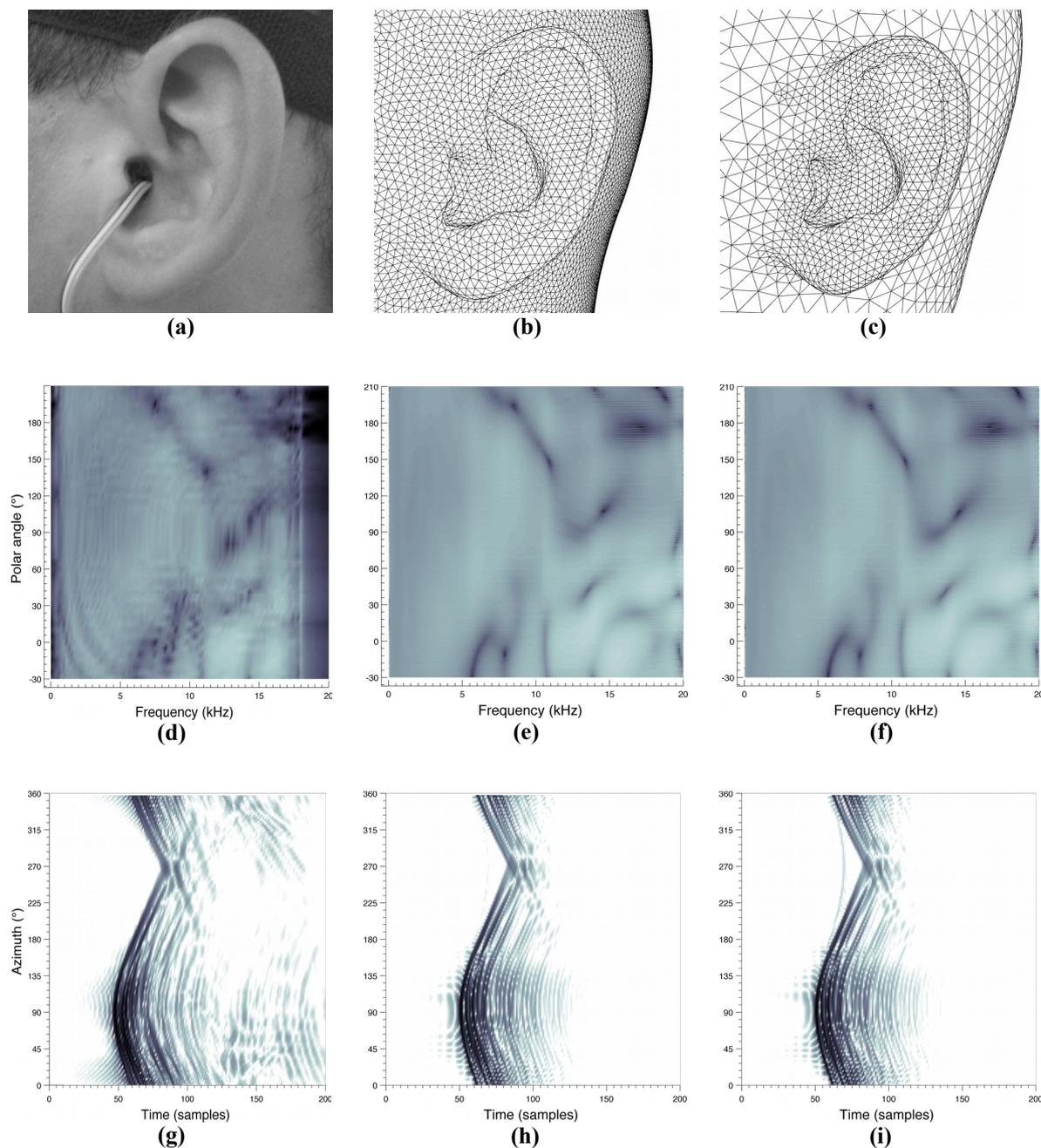


Figure 3. Meshes (a-c), amplitude DTF spectra for the median plane (d-f), and energy time curves for the horizontal plane (g-i). The polar angle of 0° , 90° , 180° corresponds to the front, top, and back position of the sound source. The azimuth of 0° , 90° , 180° and 270° corresponds to the front, left, back and right position of the sound source. The brightness encodes the magnitude in dB. (a, d, g) Acoustically measurement. (b, e, h) Uniformly sampled geometry. (c, f, i) Non-uniformly sampled geometry.

exemplary non-uniformly sampled geometry are shown in Table 1. HRTFs were calculated for 16 022 directions at a distance of 1.2 m. The resolution was 2° for both lateral and polar angles. Finally, DTFs were calculated following the same procedure as for the acoustically measured HRTFs.

Figure 3 (d, e, f) shows the amplitude spectra of DTFs in the median-plane and Figure 3 (g, h, i) shows the ETCs of HRTFs in the horizontal plane. From the visual inspection, the measured and both calculated DTFs seem to be similar. The modeled sound-localization performance (see Table 1) shows further evidence for the similarity between the HRTFs based on the uniformly and non-uniformly sampled geometry.

Figure 4 shows the relative computation time (a) for solving Eq. 5 compared to the uniform mesh and the corresponding modeled sound-localization performance parameters (b, c) as functions of the maximal edge length in the meshes. Further, the predicted relative computation time $t = \frac{N_2 \log(N_2)}{N_1 \log(N_1)}$ is shown, where N_1 is the number of elements of the uniform mesh and N_2 the number of elements of the corresponding non-uniform mesh.

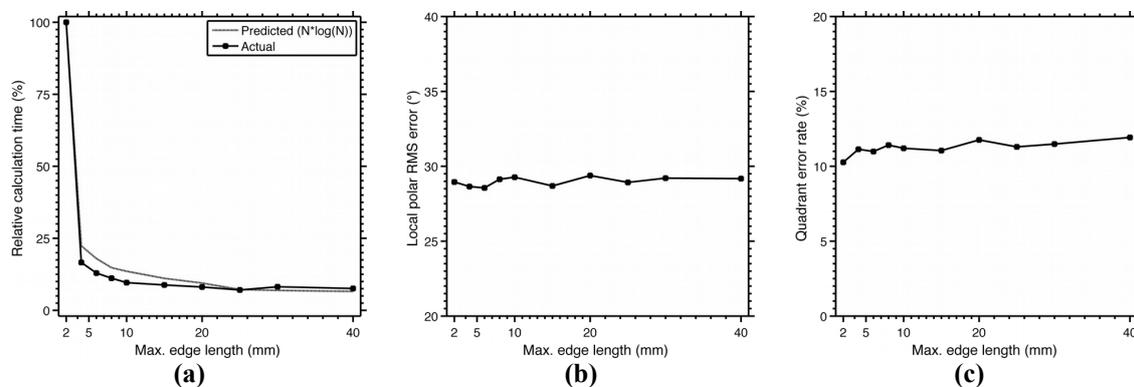


Figure 4. Relative computation time in percent as function of the maximal edge length in the mesh (a) and modeled sound-localization performance parameters (b, c).

5. Discussion

The results of our experiments with a non-uniform sampling scheme allow two observations. First, no significant decrease of modeled sound-localization performance was observed when the non-uniform sampling scheme was applied and the maximal edge length at the contra-lateral side was increased. Thus, our experiments seem to confirm our hypothesis. Second, by applying our non-uniform sampling scheme with a minimal edge length of 1 mm on the pinna and a maximal edge length of 25 mm at the contra-lateral head side the computation time has been decreased by 93 percent.

6. Conclusions

In this study, the acoustically measured HRTFs were compared to HRTFs calculated for uniformly (over 98k elements) and non-uniformly sampled (down to 9k elements) meshes. The results show similar amplitude spectra and modeled localization performance, verifying our assumptions and indicating that the generally applied recommendation of at least six elements per wavelength can be loosened for larger distances from the modeled sound source. Our non-uniform sampling scheme reduced the calculation time significantly. These results appear to be very promising for a more deeply investigation of the non-uniform sampling in the calculation of HRTFs.

7. Acknowledgements

We thank Andreas Reichinger (VRVis, Vienna) and his team (Christa Rothner, Florian Rist, Daniel Habe, and Matthias Kucera) for providing the geometry. This study was supported by the Austrian Science Fund (FWF, P 24124-N13).

REFERENCES

- ¹ Blauert, J., *Spatial hearing: the psychophysics of human sound localization*, MIT Press, Cambridge, MA, 1–494, (1997).
- ² Møller, H., Fundamentals of binaural technology, *Appl. Acoust.*, **36**, 171–218, (1992).
- ³ Majdak, P., Goupell, M. J. & Laback, B., 3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training, *Atten. Percept. Psychophys.*, **72**, 454–469, (2010).
- ⁴ Møller, H., Sørensen, M. F., Hammershøi, D. & Jensen, C. B., Head-Related Transfer Functions of Human Subjects, *J. Audio Eng. Soc.*, **43**, 300–321, (1995).
- ⁵ Katz, B. F. G., Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation, *J. Acoust. Soc. Am.*, **110**, 2440–2448, (2001).
- ⁶ Katz, B. F. G., Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements, *J. Acoust. Soc. Am.*, **110**, 2449–2455, (2001).
- ⁷ Kahana, Y. & Nelson, P. A., Numerical modelling of the spatial acoustic response of the human pinna, *Journal of Sound and Vibration*, **292**, 148–178, (2006).
- ⁸ Kahana, Y. & Nelson, P. A., Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models, *Journal of Sound and Vibration*, **300**, 552–579, (2007).
- ⁹ Gumerov, N. A., O’Donovan, A. E., Duraiswami, R. & Zotkin, D. N., Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation, *J. Acoust. Soc. Am.*, **127**, 370–386, (2010).
- ¹⁰ Kreuzer, W., Majdak, P. & Chen, Z., Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range, *J. Acoust. Soc. Am.*, **126**, 1280–1290, (2009).
- ¹¹ Marburg, S., Six boundary elements per wavelength: is that enough?, *Journal of Computational Acoustics*, **10**, 25–51, (2002).
- ¹² Morse, P. M. & Ingard, K. U., *Theoretical acoustics*, Princeton University Press, Princeton, NJ, 1-949, (1986).
- ¹³ Baumgartner, R., Majdak, P. & Laback, B. in *The Technology of Binaural Listening* (Springer, 2013), ISBN: 978-3-642-37762-4.
- ¹⁴ Burton, A. J. & Miller, G. F., The Application of Integral Equation Methods to the Numerical Solution of Some Exterior Boundary-Value Problems, *Proc. R. Soc. Lond. A*, **323**, 201–210, (1971).
- ¹⁵ Hunter, P. & Pullan, A., *FEM/BEM Notes*, (2002).
- ¹⁶ Fischer, M., Gauger, U. & Gaul, L., A multipole Galerkin boundary element method for acoustics, *Engineering Analysis with Boundary Elements*, **28**, 155–162, (2004).
- ¹⁷ Chen, Z.-S., Waubke, H. & Kreuzer, W., A formulation of the fast multipole boundary element method for acoustic radiation and scattering from three-dimensional structures, *Journal of Computational Acoustics*, **16**, 303–320, (2008).
- ¹⁸ Lorensen, W. E. & Cline, H. E., Marching Cubes: A High Resolution 3D Surface Construction Algorithm, in *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, ACM, 163–169, (1987).
- ¹⁹ Ziegelwanger, H., Reichinger, A. & Majdak, P., Calculation of listener-specific head-related transfer functions: Effect of mesh quality, *Proceedings of the 21st International Congress on Acoustics*, Montreal, **19**, 050017, (2013).
- ²⁰ Möbius, J. & Kobbelt, L., OpenFlipper: An Open Source Geometry Processing and Rendering Framework, *Proceedings of the 7th International Conference on Curves and Surfaces*, Springer-Verlag, 488–500, (2012).
- ²¹ Majdak, P., Balazs, P. & Laback, B., Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions, *J. Audio Eng. Soc.*, **55**, 623–637, (2007).
- ²² Middlebrooks, J. C. & Green, D. M., Directional dependence of interaural envelope delays, *J. Acoust. Soc. Am.*, **87**, 2149–2162, (1990).
- ²³ Ziegelwanger, H. & Majdak, P., Modeling the direction-continuous time-of-arrival in head-related transfer functions, *J. Acoust. Soc. Am.*, **135**, 1278-1293, (2014).

Chapter 7

A-priori mesh grading for the numerical calculation of the HRTFs

On the 20th February 2016, this work was submitted to Applied Acoustics as:

Ziegelwanger, H., Kreuzer, W., and Majdak, P. (**submitted**). "A-priori mesh grading for the numerical calculation of the head-related transfer functions," *Applied Acoustics*.

The idea behind the a-priori mesh grading came from me, the first author. I designed and implemented the algorithm based on an existing code for the uniform re-meshing of geometry. Together with the second author, I designed the evaluation of the method. I numerically calculated HRTFs, evaluated the method, generated all figures and wrote the initial draft of the manuscript with the help of the second author. The third author helped writing the manuscript.

Co-author 1:

Co-author 2:

A-priori mesh grading for the numerical calculation of the head-related transfer functions

Harald Ziegelwanger^{a,*}, Wolfgang Kreuzer^b, Piotr Majdak^b

^aAIT Austrian Institute of Technology GmbH, Mobility Department, Transportation Infrastructure Technologies, Giefinggasse 2, 1210 Vienna, Austria

^bAcoustics Research Institute, Austrian Academy of Sciences, Wohllebengasse 12-14, Vienna, 1040, Austria

Abstract

Head-related transfer functions (HRTFs) describe the directional filtering of the incoming sound caused by the morphology of a listener's head and pinnae. When an accurate model of a listener's morphology exists, HRTFs can be calculated numerically with the boundary element method (BEM). However, the general recommendation to model the head and pinnae with at least six elements per wavelength renders the BEM as a time-consuming procedure when calculating HRTFs for the full audible frequency range. In this study, a mesh preprocessing algorithm is proposed, viz., a-priori mesh grading, which reduces the computational costs in the HRTF calculation process significantly. The mesh grading algorithm deliberately violates the recommendation of at least six elements per wavelength in certain regions of the head and pinnae and varies the size of elements gradually according to an a-priori defined grading function. The evaluation of the algorithm involved HRTFs calculated for various geometric objects including meshes of three human listeners and various grading functions. The numerical accuracy and the predicted sound-localization performance of calculated HRTFs were analyzed. A-priori mesh grading appeared to be suitable for the numerical calculation of HRTFs in the full audible frequency range and outperformed uniform meshes in terms of numerical errors, perception based predictions of sound-localization performance, and computational costs.

Keywords: Head-related transfer functions, Boundary element method, Mesh grading

PACS: 43.66.Qp, 43.66.Pn, 43.64.Ha, 43.20.Fn

1. Introduction

The head-related transfer functions (HRTFs) describe the directional filtering of incident sound waves at the entrance of a listener's ear-canal (Mehrgardt and Mellert, 1977; Wightman and Kistler, 1989). This filtering is caused by the overall diffraction, shadowing, and reflections at human anatomical structures, i.e., the torso, head, and pinnae. Thus, HRTFs are closely related to a listener's individual geometry and they provide listener-specific spectral (Middlebrooks, 1999) and temporal features (Macpherson and Middlebrooks, 2002) which are essential for three-dimensional (3D) sound localization, e.g., in binaural audio reproduction systems (Møller, 1992).

HRTFs are usually acquired acoustically in a resource-demanding process, in which small microphones are placed into listener's ear canals and transfer functions are measured for many directions in an anechoic chamber (Møller et al., 1995; Majdak et al., 2010). HRTFs can also be acquired by means of a numerical HRTF calculation, i.e., by simulating the sound field of an incident wave scattered by a listener's head and pinnae. In recent years, the boundary element method (BEM, Gaul et al., 2003) became a powerful tool for such simulations in acoustics.

The BEM was used in many studies for the numerical calculation of HRTFs (Katz, 2001a,b; Kahana and Nelson, 2006, 2007; Gumerov et al., 2010; Kreuzer et al., 2009; Rui et al., 2013; Jin et al., 2014; Ziegelwanger et al., 2015b). In general, the numerical HRTF calculation is based on a 3D polygon mesh, i.e., a set of nodes and elements with these nodes as vertices, describing a listener's morphology.

In element-based acoustic simulations resolution of the mesh should be related to the wavelength of the simulated frequency (Marburg, 2002). The mesh resolution is measured by the number of elements per wavelength or by the average length of edges in the mesh, i.e., the average edge length (AEL, Ziegelwanger et al., 2015b). The accuracy of the numerical calculations depends on these metrics. In Marburg (2002) the relative numerical error was below fifteen percent, when at least six elements per wavelength were used. Gumerov et al. (2010) recommended five elements per wavelength, equilateral triangles, and a valence of six, i.e., the number of edges incident to a vertex describing the regularity of a mesh, and a uniform vertex distribution in the mesh. In Ziegelwanger et al. (2015b), an AEL of 1 mm to 2 mm was required for accurate numerical HRTF calculations. Given the average human body surface area and a frequency range of up to 18 kHz, these recommendations result in a uniform head and pinna mesh consisting of approximately 100 000 equilateral triangular elements.

The computational costs of the BEM, i.e., the calculation time and the required amount of physical memory, are gen-

*Corresponding author. Tel.: +43 50550-6611.

Email addresses: harald.ziegelwanger@ait.ac.at (Harald Ziegelwanger), wolfgang.kreuzer@oeaw.ac.at (Wolfgang Kreuzer), piotr@majdak.com (Piotr Majdak)

erally high and increase with the number of elements in the mesh (Kreuzer et al., 2009). The first numerical HRTF calculations were limited to 22 000 elements and frequencies up to 5.4 kHz because the calculation time was in the range of tens of hours for a single frequency (Katz, 2001a). The HRTF calculation became feasible for the full audible frequency range (Kreuzer et al., 2009; Gumerov et al., 2010) by coupling the BEM with the fast multipole method (FMM, e.g., Chen et al., 2008). HRTFs calculated with the FMM showed good results for artificial heads by means of visual comparison of amplitude spectra (Gumerov et al., 2010) and for human listeners by means of individual sound-localization performance (Ziegelwanger et al., 2015b). However, the numerical HRTF calculation process for the full audible frequency range can still last tens of hours on a single CPU (Kreuzer et al., 2009; Ziegelwanger et al., 2015b).

While the computational costs can be reduced by reducing the number of elements in the mesh, a simple mesh coarsening considering all elements, i.e., a *uniform* re-meshing, also reduces the accuracy of the numerical calculation (Marburg, 2002). The loss of accuracy is caused by geometric and numerical error (Treeby and Pan, 2009). The geometric error arises because of under-sampling the geometry and the numerical error arises because of under-sampling the sound field on the geometry (Ziegelwanger et al., 2015b). In other fields of computational physics, more sophisticated geometry discretization methods, resulting in *non-uniform* meshes, have been investigated. For the finite-element method, the numerical error introduced by goal-oriented mesh adaptivity (Walsh and Demkowicz, 2003), non-uniform meshes (Goldstein, 1982), and mesh grading (Heinrich et al., 1996) were investigated. For elliptic boundary-value problems, *a-priori mesh grading* was proposed (Langer et al., 2015), where the element size was varied based on a-priori knowledge of the location of singular points, i.e., sharp edges and corners in the geometry or discontinuities in the boundary conditions. For the 2D-BEM and hyperbolic boundary-value problems, adaptive meshes were investigated (Chen et al., 2002; Liang et al., 1999). In general, non-uniform meshes showed better convergence rates than uniform meshes and the accuracy was higher for non-uniform meshes than for uniform meshes. Even though the investigations for hyperbolic boundary-value problems were done for the 2D-BEM only, a non-uniform mesh of a human head seems to be a promising approach to reduce the computational costs in the numerical HRTF calculation process.

Hence, in this article, we adapt the idea of a-priori mesh grading for the numerical calculation of HRTFs. First, we briefly review the BEM (Sec. 2) and describe our a-priori mesh grading algorithm (Sec. 3). In Sec. 4, we show its evaluation. Secs. 4.1 and 4.2 describe the software and metrics we have used for the evaluation. Sec. 4.3 and 4.4, show the evaluation of various grading functions based on a comparison to reference HRTFs of a sphere. Then, the most promising grading functions were evaluated on the geometry of a pinna (Sec. 4.5). Finally, in Sec. 4.6, the best performing grading function was applied on meshes of human heads, for which the HRTFs were evaluated by means of numerical and perceptual errors.

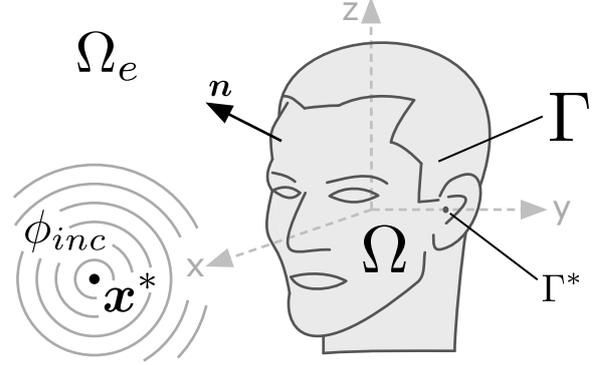


Figure 1: Schematic representation of the exterior scattering problem for the numerical calculation of HRTFs. A point source is placed at \mathbf{x}^* and emits the incident sound field $\phi_{inc}(\mathbf{x})$ in Ω_e outside a listener's head Ω with surface Γ and the microphone area Γ^* . x , y , and z represent the Cartesian coordinate system as described in Ziegelwanger and Majdak (2014).

2. Boundary element method

The boundary element method in the context of calculating HRTFs is schematically shown in Fig. 1. Here, the object Ω with boundary Γ represents the scatterer, i.e. the human head and pinnae. Ω_e defines the domain outside the scatterer. A point source at \mathbf{x}^* (in the following called *loudspeaker position*) emits spherical waves, i.e., it produces the incident sound field $\phi_{inc}(\mathbf{x})$. Γ^* is the *microphone area* at the entrance of the ear canal.

The total sound field $\phi(\mathbf{x})$ is described by the boundary integral equation:

$$\tau\phi(\mathbf{x}) = \int_{\Gamma} H(\mathbf{x}, \mathbf{y})\phi(\mathbf{y})d\mathbf{y} - \int_{\Gamma} G(\mathbf{x}, \mathbf{y})v(\mathbf{y})d\mathbf{y} + \phi_{inc}(\mathbf{x}), \quad (1)$$

where $G(\mathbf{x}, \mathbf{y})$ and $H(\mathbf{x}, \mathbf{y})$ are the Green's function of the Helmholtz equation and its derivative with respect to the normal vector \mathbf{n} to Γ at the point \mathbf{y} . $\phi(\mathbf{x}) = \frac{p(\mathbf{x})}{i\omega\rho}$, $p(\mathbf{x})$ and $v(\mathbf{x}) = \frac{\partial\phi(\mathbf{x})}{\partial n} = \mathbf{n} \cdot \nabla\phi(\mathbf{x})$ denote the velocity potential, the sound pressure, and the particle velocity at a point \mathbf{x} , respectively. ρ denotes the density of air and τ is a factor depending on the position of \mathbf{x} . $\tau = 1$ for $\mathbf{x} \in \Omega_e$, $\tau = 1/2$ for $\mathbf{x} \in \Gamma$, and $\tau = 0$ when \mathbf{x} is located inside Ω . The scatterer is assumed to be rigid, thus $\frac{\partial\phi(\mathbf{x})}{\partial n} = 0$ for $\mathbf{x} \in \Gamma$.

To speed up calculations, HRTFs are determined by applying the principle of reciprocity (Morse and Ingard, 1986), where the roles of sources and receivers are exchanged. To this end, the in-ear microphone is simulated by a point source close to Γ^* (Gumerov et al., 2010) or by active vibrating elements at Γ^* (Ziegelwanger et al., 2015b; Kreuzer et al., 2009). HRTFs are evaluated using the calculated sound pressure at the loudspeaker positions. For the active vibrating elements, this means technically that the contribution of an external sound source $\phi_{inc}(\mathbf{x})$ is replaced by an additional boundary condition $v(\mathbf{x}) \neq 0$ for $\mathbf{x} \in \Gamma^*$.

In our approach, HRTFs are calculated numerically in three steps. First, Γ is discretized as a 3D polygon mesh \mathcal{M} , consisting of vertices \mathcal{V} , edges \mathcal{E} and elements \mathcal{F} (Fig. 3a), and the unknown solution $\phi(\mathbf{x})$ on Γ is approximated using simple basis functions (Hunter and Pullan, 2002), e.g., piecewise constant basis functions. Using a collocation approach Eq. 1 (for $\tau = 0.5$) is transformed into a linear system of equations $\mathbf{A}\phi = \mathbf{b}$, where ϕ is the vector of unknown velocity potentials. The Burton-Miller approach (Burton and Miller, 1971) is used to ensure a unique solution. For details about the derivation of the stiffness matrix \mathbf{A} and the right-hand-side \mathbf{b} refer to Chen et al. (2008) and Ziegelwanger et al. (2015a). Second, the solution for the linear system of equations is calculated by using an iterative solver. The FMM is used to speed up the matrix-vector multiplications needed for the iterative solver (Chen et al., 2008). Third, given the solution ϕ at the boundary Γ , the sound pressure $p(\mathbf{x}) = i\rho\omega\phi(\mathbf{x})$ at any point \mathbf{x} in the exterior domain Ω_e , e.g., the loudspeaker grid, is determined by evaluating Eq. 1 (for $\tau = 1$).

3. A-priori mesh grading

The a-priori mesh grading approach aims at reducing the number of elements $\#\mathcal{F}$ while preserving the accuracy in the calculation results by gradually increasing the length of edges in \mathcal{M} as a function of the distance of an edge to Γ^* . The validity of our approach is based on two assumptions.

As for the geometric error, we assume that the geometry of the ipsilateral pinna¹ has the main influence on the accuracy of the HRTFs (see also Ziegelwanger et al., 2015b), whereas the geometry of the rest of the head (including the contralateral pinna) plays a minor role. Thus, the effect of a geometric error caused by large elements at the contralateral head side is assumed to be negligible.

As for the numerical error, we assume that the discretization of $\phi(\mathbf{x})$ in the proximity range of a discontinuity in boundary conditions, i.e., the jump in the neumann boundary condition at Γ^* in the reciprocal calculation, has to be very fine, whereas the effect of the size of elements in regions far apart from Γ^* is negligible, because the Green's function $G(\mathbf{x}, \mathbf{y})$ and its derivatives in Eq. 1 decay fast with increasing distance $\|\mathbf{x} - \mathbf{y}\|$. Additionally the amplitude of $\phi(\mathbf{x})$ on the contralateral side will be much smaller than on the ipsilateral side (with a difference of approximately 90 dB). Therefore, any error caused by the increased element size at the contralateral side will be damped in the order of tens of dB, introducing only small artifacts to the final HRTFs.

Based on our assumptions, an algorithm which gradually increases the length of edges in \mathcal{M} as a function of the distance of an edge to Γ^* was designed. The algorithm is based on a re-meshing algorithm for uniform meshes (Botsch and Kobbelt, 2004). In this algorithm, if the length of an edge does not match

¹When using the reciprocity approach, HRTFs need to be calculated for both ears separately. Thus, in this context, the terms ipsilateral and contralateral correspond to the side of the head, where the sound source is located or not, respectively.

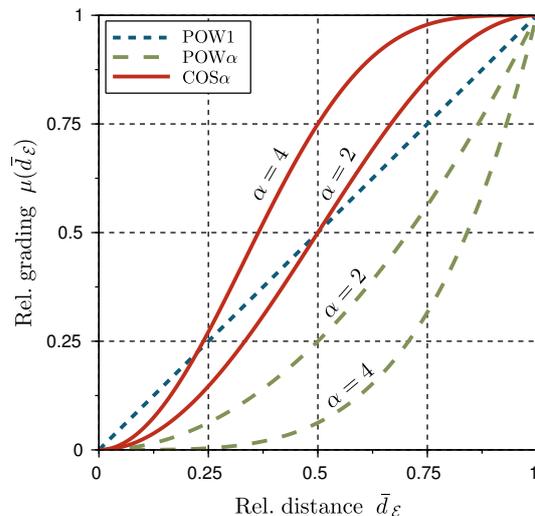


Figure 2: Relative grading resulting from various grading functions (linear POW1, higher-order grading POW α , and raised-cosine grading COS α).

a pre-defined target edge length, the mesh around this edge is modified. In our approach, this algorithm is extended such that the target edge length is described by a grading function dependent on the (relative) distance of an edge to Γ^* . In the following, we describe the details of the grading functions and of the re-meshing algorithm.

3.1. Grading functions

We consider the relative distance of an edge \mathcal{E} to Γ^* :

$$\bar{d}_{\mathcal{E}} = \frac{|\mathcal{E}_m - \Gamma_m^*|}{d_{max}}, \quad (2)$$

where \mathcal{E}_m is the midpoint of \mathcal{E} , Γ_m^* is the midpoint of Γ^* , and $d_{max} = \max_{\mathcal{E}} d_{\mathcal{E}}$ is the maximum distance of edges in \mathcal{M} . Note that $\bar{d}_{\mathcal{E}} = 0$ and $\bar{d}_{\mathcal{E}} = 1$ correspond to the most ipsilateral and most contralateral side of the head, respectively.

For the grading of \mathcal{M} , we propose various grading functions $\mu(\bar{d}_{\mathcal{E}})$. These functions, shown in Fig. 2, can be structured in two classes. The first class consists of power functions:

$$\mu(\bar{d}_{\mathcal{E}}) = \bar{d}_{\mathcal{E}}^{\alpha}, \quad (3)$$

with $\alpha = 1$ as the first-order grading (**POW1**), which increases the length of edges linearly from the ipsilateral to the contralateral head side. Also, higher-order grading functions were considered, e.g., quadratic grading (**POW2**) and quartic grading (**POW4**) for $\alpha = 2$ and $\alpha = 4$, respectively. With increasing α , these grading functions increase the element size slightly at the ipsilateral side and rapidly at the contralateral side. Note that a zeroth-order grading yields a uniform mesh.

The second class is based on raised cosine:

$$\mu(\bar{d}_{\mathcal{E}}) = 1 - \cos^{\alpha}(\pi\bar{d}_{\mathcal{E}}/2), \quad (4)$$

with the second-order cosine grading (**COS2**) for $\alpha = 2$ and fourth-order cosine grading (**COS4**) for $\alpha = 4$. In contrast to the $\text{POW}\alpha$ functions, $\text{COS}\alpha$ functions concentrate small edges around Γ^* . While $\alpha = 2$ then increases the edge length almost linearly from the ipsilateral to the contralateral head side, $\alpha = 4$ grading increases the edge length rapidly to the maximum edge length.

In order to calculate the target edge length for each edge in \mathcal{M} , two global mesh grading parameters are defined, the minimum target edge length $\hat{\ell}_{\min}$ and the maximum target edge length $\hat{\ell}_{\max}$. The local target edge length $\hat{\ell}_{\bar{d}_E}$ is then calculated as:

$$\hat{\ell}_{\bar{d}_E} = \hat{\ell}_{\min} + (\hat{\ell}_{\max} - \hat{\ell}_{\min})\mu(\bar{d}_E). \quad (5)$$

3.2. Re-meshing algorithm

Most of the re-meshing algorithms in the field of computer graphics use triangle mesh decimation, vertex clustering, or voxel based object simplification (Luebke, 2001). Our a-priori mesh grading algorithm is based on an efficient re-meshing algorithm from Botsch and Kobbelt (2004)². This algorithm modifies a mesh in an iterative procedure with the goal to obtain a uniform mesh with a given target edge length. In our mesh grading algorithm, the target edge length is the local target edge length $\hat{\ell}_{\bar{d}_E}$ from the previous section. Thus, in each iteration, first, edges are split if $\ell_E > \frac{4}{3}\hat{\ell}_{\bar{d}_E}$ (Fig. 3b). Second, edges are collapsed if $\ell_E < \frac{4}{5}\hat{\ell}_{\bar{d}_E}$ (Fig. 3c). Third, edges are flipped if the valence of neighboring vertices is larger than six (Fig. 3d). Last, vertices are relocated on the surface of the original mesh by tangential smoothing (Botsch and Kobbelt, 2004). In our study, ten iterations were sufficient to achieve the edge length distribution targeted by the corresponding grading function.

4. Evaluation

The effect of a-priori mesh grading on numerically calculated HRTFs was investigated in three steps, with different geometric objects in each step. First, HRTFs were numerically calculated for a rigid sphere (**SPH**) object and were compared to an analytical solution. Second, the same HRTFs were compared to HRTFs calculated for a uniform high-resolution mesh of that sphere, validating the usability of a high-resolution mesh to generate a reference HRTF set for further investigations. Third, the SPH object was extended by a generic pinna yielding the sphere-and-pinna (**SAP**) object. For that object, an analytical solution is not available and HRTFs numerically calculated from a high-resolution mesh were used as the reference. Fourth, HRTFs were calculated for three human objects (**HUMs**) represented by head and pinna models of actual human listeners. For the HUMs objects, the effect of a-priori mesh grading on numerically calculated HRTFs was (in addition to numerical

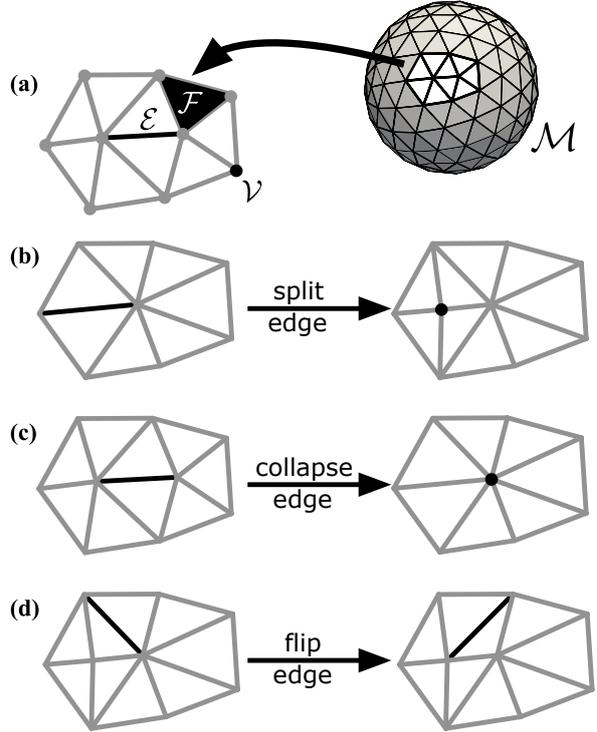


Figure 3: The re-meshing algorithm. (a) Definitions: \mathcal{V} , \mathcal{E} , and \mathcal{F} represent a vertex, an edge, and a face, respectively. Mesh operations: (b) edge splitting, (c) edge collapsing, and (d) edge flipping.

errors) evaluated from the perceptual point of view. In this section, we first describe some general aspects of our evaluation and then show and discuss the results for the evaluated objects.

4.1. HRTF calculation

Mesh2HRTF³ was used to calculate HRTFs numerically for 200 frequencies, which were linearly spaced between 0.1 kHz and 20 kHz. Calculations were run on a Linux cluster consisting of eight PCs with Intel i7-3820 processors running at 3.6 GHz. Each PC was equipped with 64 GB of RAM. In total, more than hundred HRTF sets were calculated in this study and the calculations lasted several days.

HRTF positions were defined by means of two loudspeaker grids. A loudspeaker grid represents the evaluation nodes in the BEM. In the first grid, the equi-angular (**EQA**) grid, the interaural-polar coordinate system was used (Morimoto and Aokata, 1984). In that system, the poles are placed along the interaural axis and a direction is represented by lateral angle ϑ and polar angle φ . In that grid, 5042 virtual loudspeakers were placed equi-angular on a sphere with radius 1.2 m. The resolution was 2.5° and 5° in the lateral and polar dimension,

²Available in OpenFlipper, version 2.1, (Möbius and Kobbelt, 2012): <http://www.openflipper.org> (date last viewed: January 31, 2016)

³Mesh2HRTF: version 0.1.2, available from <http://mesh2hrtf.sourceforge.net> (date last viewed: January 31, 2016)

respectively. The EQA grid was used to analyze the *numerical* accuracy of the numerically calculated HRTFs.

In the second grid, the **ARI** grid, 1550 virtual loudspeakers were placed on a sphere with radius 1.2 m (Majdak et al., 2010; Ziegelwanger et al., 2015b). The spherical coordinate system was used which is described by the azimuth and elevation angle. The grid had a polar gap at elevation angles below -30° , as well as a resolution of 5° in the elevation and 2.5° and 5° for azimuth angles smaller and larger than 30° , respectively. The ARI grid was used in the *perceptual* evaluation of the numerically calculated HRTFs.

4.2. Considered metrics

Two error metrics were used in the evaluation. First, the accuracy of numerical calculation was considered which shows the performance and limits of the method and frames our results into context of previous studies in the field of computational acoustics. Second, individual sound-localization performance was considered which shows the perceptual impact of the numerical error and the relevance of the method in the context of HRTFs and frames our research into context of previous studies in the field of spatial hearing.

The accuracy of the numerical calculation was quantified by the relative numerical error. To this end, the complex-valued sound pressure from a numerical calculation was compared with a reference solution resulting in the relative numerical error $e_{LP}^{\Omega_e}$ given by:

$$e_{LP}^{\Omega_e} = \frac{\|p_{num}^{\Omega_e} - p_{ref}^{\Omega_e}\|_{L^p}}{\|p_{ref}^{\Omega_e}\|_{L^p}}, \quad (6)$$

where p_{num} and p_{ref} are the evaluated and a reference sound pressure, respectively. $\|\cdot\|_{L^p}$ denotes an L^p -norm.

Two L^p -norms were considered, viz., the L^2 -norm and the L^∞ -norm. For the calculated frequencies and the EQA grid, the L^2 norm was calculated as:

$$\|f\|_{L^2} \approx \sqrt{\Delta\omega\Delta\varphi\Delta\vartheta \sum_i \sum_j \sum_k |f(\omega_i, \varphi_j, \vartheta_k)|^2 \sin(\vartheta_k)}, \quad (7)$$

where (φ_j, ϑ_k) are defined by the EQA grid. The L^∞ -norm was:

$$\|f\|_{L^\infty} = \max_{\omega, \varphi, \vartheta} |f(\omega, \varphi, \vartheta)|. \quad (8)$$

Note that our error formulation in Eq. 6 does not separate magnitude and phase differences and thus considers both. A phase difference of π results in a relative error of two hundred percent.

The individual sound-localization performance was quantified in terms of model predictions for an HRTF set. Two models were used. First, spectral features were analyzed with a model of sound-localization performance in sagittal-planes (Baumgartner et al., 2013) implemented as `baumgartner2013` in the Auditory Modeling Toolbox⁴ (AMT, S ndergaard and

Majdak, 2013). This model predicts the individual sound-localization performance by means of the local polar RMS error (PE) and the quadrant error rate (QE), which are common error metrics for analyzing results of sound-localization experiments Middlebrooks (1999). Second, temporal features were analyzed by a direction-continuous model of the time-of-arrival (TOA, Ziegelwanger and Majdak, 2014) implemented as `ziegelwanger2014` in the AMT. In the TOA model, the temporal features are quantified as the equivalent head radius resulting from the tested HRTF set. A deviation from the radius obtained for a reference HRTF set can be interpreted as an artifact in the broadband temporal properties of tested HRTFs, having potentially an effect on the perception of the interaural time differences with such HRTFs.

4.3. The SPH object and analytical reference

The SPH object was a rigid sphere model with a radius R of 100 mm. First, a high-resolution mesh of the sphere was constructed in Blender⁵. It consisted of 139 194 elements corresponding to a resolution of about 14 elements per wave length at 18 kHz and an AEL of approximately 1.4 mm. Then, by uniformly re-meshing that mesh, uniform meshes (UNI) were created with AELs ranging from 2 mm to 20 mm. Further, by applying the a-priori mesh grading to the high-resolution mesh, graded meshes were created with $\#\mathcal{F}$ s and AELs covering the range of the uniform meshes. Five grading functions were considered (POW1, POW2, POW4, COS2, and COS4). Table 1 shows the relevant mesh and grading parameters for all tested conditions.

A point source was placed on the y-axis close to the surface of the sphere $\mathbf{x}^* = [0, 101, 0]$ mm. The EQA grid was used as loudspeaker grid.

The relative numerical error (see Eq. 6) was calculated. The reference was an analytically derived HRTF set of the SPH object. To this end, the sound field of a point source scattered by a rigid sphere (Beranek and Mellow, 2012) was calculated by:

$$p_{ref}(\mathbf{x}) = p_{inc}(\mathbf{x}) + p_{scat}(\mathbf{x}), \quad (9)$$

where $p_{inc}(\mathbf{x}) = p_0 \frac{e^{-ik|\mathbf{x}-\mathbf{x}^*|}}{4\pi|\mathbf{x}-\mathbf{x}^*|}$ is the incoming sound field, p_0 is the source strength, k is the wavenumber, and $|\mathbf{x}-\mathbf{x}^*|$ is the distance between the loudspeaker position and the source. The scattered sound field p_{scat} caused by the sphere is then:

$$p_{scat}(\mathbf{x}) = \frac{ikp_0}{4\pi} \sum_{n=0}^{\infty} (2n+1)h_n^{(2)}(k|\mathbf{x}^*|) \frac{j'_n(kR)}{h_n^{(2)}(kR)} h_n^{(2)}(k|\mathbf{x}|) P_n(\cos\beta), \quad (10)$$

where j_n and $h_n^{(2)}$ are the spherical Bessel and (second order) Hankel functions, respectively, and j'_n and $h_n^{(2)}$ are their derivatives. P_n are the Legendre polynomials of order n . $\cos(\beta) = (\mathbf{x}^T \mathbf{x}^*) / (|\mathbf{x}||\mathbf{x}^*|)$ represents the angle between the source and the

⁴AMT: version 0.9.6, available from <http://amttoolbox.sourceforge.net> (date last viewed: January 31, 2016)

⁵Blender: version 2.72b, available from <http://www.blender.org> (date last viewed: January 31, 2016)

	$\hat{\ell}$ (target)		ℓ (actual)			$\#\mathcal{F}$
	$\hat{\ell}_{min}$	$\hat{\ell}_{max}$	ℓ_{min}	ℓ_{max}	ℓ_{avg}	
UNI	2		1.4	2.1	1.9	81920
	3		1.7	4.4	2.9	35490
	4		2.5	5.3	3.8	20138
	5		3.0	6.7	4.9	12358
	10		5.5	13.3	9.4	3278
POW1	2	5.1	1.4	5.2	2.8	35798
	2	7.5	1.4	7.7	3.6	20002
	2	10.5	1.5	10.5	4.6	12174
	2	16	1.6	15.0	6.1	6432
	2	50	1.8	42.8	14.0	994
POW2	2	7.2	1.2	7.3	2.7	35498
	2	12.5	1.4	11.9	3.3	20270
	2	20	1.4	19.2	3.9	12550
	2	40	1.5	34.8	4.8	6286
	8	60	3.9	54.1	13.9	1068
POW4	2	20	1.1	18.6	2.6	29934
	2	40	1.1	31.1	2.8	20722
	4	40	1.8	35.5	4.9	8160
COS2	2	5	1.1	5.2	2.8	35024
	2	11	1.0	11.4	4.1	13474
	2	20	1.5	20.1	8.8	2714
COS4	2	8	1.2	8.2	4.2	14420
	2	15	1.2	15.3	6.0	5924
	2	20	1.0	19.7	6.8	4088

Table 1: The SPH objects: Mesh statistics for uniform (UNI) and graded (POW α and COS α) meshes. l_{min} , l_{max} , and l_{avg} are the minimum, maximum, and average edge length in millimeter, respectively. $\#\mathcal{F}$ denotes the number of elements in a mesh.

loudspeaker position. Eq. 10 was evaluated up to the order of $n = 125$.

The relative numerical errors were calculated for the L^2 -norm (Eq. 7) and the L^∞ -norm (Eq. 8), both for all, ipsilateral, and contralateral loudspeaker positions. Further, the computation time averaged across calculated frequencies was recorded.

Fig. 4 shows the relative numerical errors $e_{L^2}^{\Omega_c}$ as functions of $\#\mathcal{F}$ calculated for various mesh conditions. $e_{L^2}^{\Omega_c}$ considered all (Fig. 4a), ipsilateral (b), and contralateral (c) loudspeaker positions. Note that the filled circle highlights the error obtained for the high-resolution mesh, a condition usually thought to provide very accurate results and evaluated as the reference in the following section.

For the uniform meshes (the UNI condition), the error increased monotonically with decreasing $\#\mathcal{F}$. Generally, for the non-uniform meshes (conditions POW α and COS α), the error also increased with decreasing $\#\mathcal{F}$. However, in conditions POW1 and POW2, the errors did not decrease much for $\#\mathcal{F}$ beyond 8000. They seem to converge at the level of the error obtained for the UNI condition with an AEL of 2 mm. In

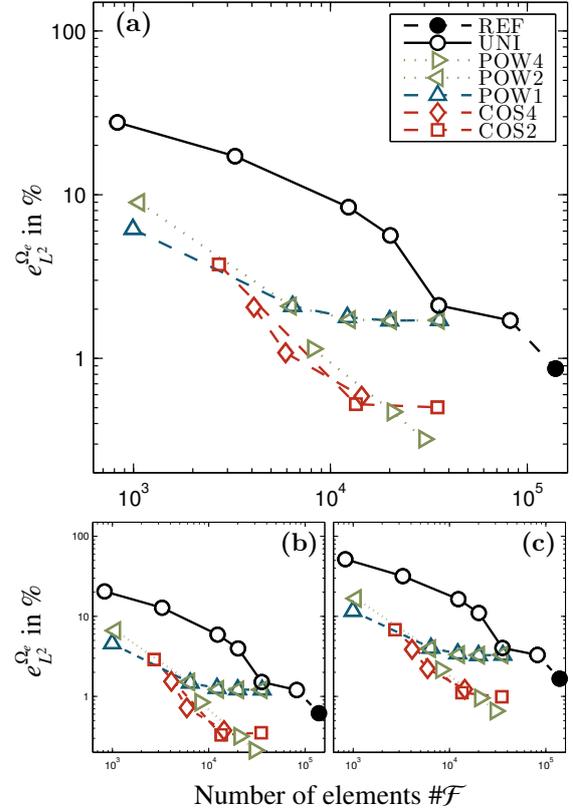


Figure 4: The SPH object and analytical reference: Relative numerical errors for (a) all nodes, (b) ipsilateral nodes, and (c) contralateral nodes of the EQA loudspeaker grid. The reference for the error was the analytically derived HRTF set following Eq. 9.

conditions COS2, COS4, and POW4, the errors decreased further even for $\#\mathcal{F}$ beyond 8000, showing a similar decay rate as the errors for the UNI condition. These grading functions showed the most promising effect of mesh grading, for which as it seems, $\#\mathcal{F}$ can be reduced by factor of ten without raising the errors significantly.

In conditions COS2, COS4, and POW4, for $\#\mathcal{F}$ s beyond 10000, the errors were even smaller than those for both the 2 mm uniform and the high-resolution meshes. This result might appear intriguing because the average resolution of the graded meshes was much smaller than the resolution of those two uniform meshes. It seems like having more elements is not always of advantage, potentially resulting from a larger numerical error when dealing with more elements in the numerical procedures of the BEM (e.g., summing over many elements of an array). In the context of our evaluation, this finding provides evidence that HRTFs based on graded meshes are able to approximate the exact analytical solution at least as good as HRTFs based on uniform meshes.

The errors for the ipsilateral loudspeakers only (Fig. 4b) were smaller than those for contralateral ones (Fig. 4c). This was expected for the graded meshes, which poorer accuracy is an

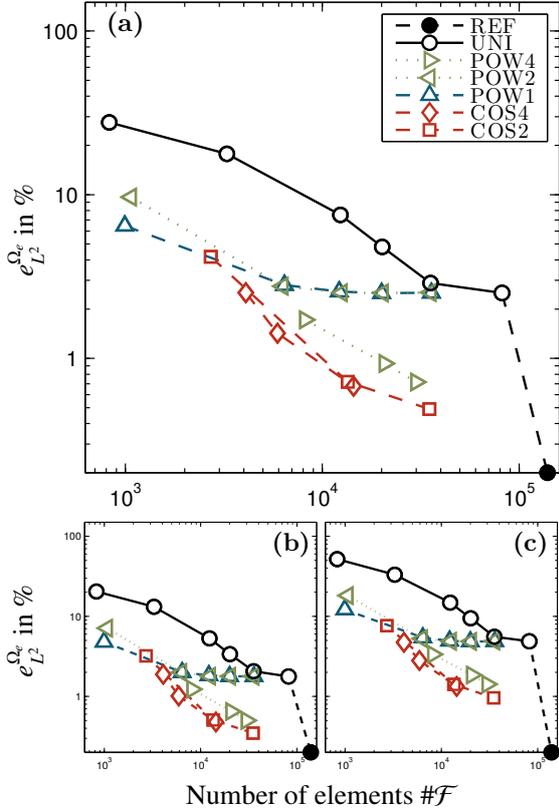


Figure 5: The SPH object and numerical reference: Relative numerical errors for (a) all nodes, (b) ipsilateral nodes, and (c) contralateral nodes of the EQA loudspeaker grid. The reference for the errors was the HRTF set numerically calculated for the high-resolution mesh.

intrinsic property for the contralateral directions. Interestingly, even in the conditions with uniform meshes, the errors for the ipsilateral loudspeakers were smaller than those for contralateral ones, which might appear surprising given similar element sizes at the two lateral sides in uniform meshes. This finding, however, shows that the elements on the ipsilateral side affect HRTFs much more than those on the contralateral side, supporting our assumption on loosening requirements for the geometric accuracy at the contralateral side.

The pattern of the relative errors based on the L^∞ -norm was similar to that based on the L^2 -norm (thus figures are not shown). Averaged across the tested conditions, $e_{L^\infty}^{\Omega_r}$ s were only 1.51 times larger than $e_{L^2}^{\Omega_r}$ s. The similarity between these two errors indicates that 1) none of the loudspeaker positions yielded an error larger than 1.51 of the average error, showing no evidence for an outlier, and 2) the same conclusions can be drawn from both norms.

4.4. The SPH object and numerical reference

For the SAP and HUM objects, an analytical derivation of the HRTFs is not feasible and an HRTF set numerically calculated from a high-resolution mesh must be used as a reference. Thus,

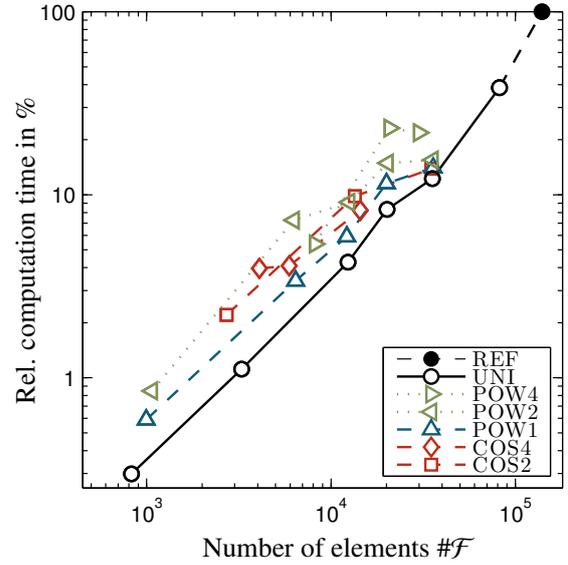


Figure 6: Relative computation time to calculate $\phi(x)$ in percent. The reference (100%) was the recorded computation time required for the high-resolution mesh.

we investigated the appropriateness of the high-resolution mesh as a basis for reference HRTFs. To this end, the relative errors from the previous section were recalculated with another reference, namely, the HRTF set calculated for a high-resolution mesh of the SPH object. Further, the gain in computation time from reducing $\#F$ in the calculations was evaluated.

Fig. 5 shows the relative numerical error $e_{L^2}^{\Omega_r}$. Note that the filled circle still highlights the error obtained for the high-resolution mesh, which is zero percent now, and is thus shown on the abscissa. The errors seem to follow similar patterns as those with analytically derived HRTFs as reference (compare Fig. 4). Pearson's correlation coefficient between $e_{L^2}^{\Omega_r}$ calculated with the analytically derived HRTFs and $e_{L^2}^{\Omega_r}$ calculated with the high-resolution mesh HRTFs was 0.99. Both the patterns and the correlation coefficient indicate that the high-resolution mesh adequately represents the reference for further investigations with more complex geometric objects for which an analytical solution is not feasible.

As for $e_{L^\infty}^{\Omega_r}$, the errors followed similar patterns as for $e_{L^2}^{\Omega_r}$. When averaged across the tested conditions, $e_{L^\infty}^{\Omega_r}$ was only 1.43 times larger than $e_{L^2}^{\Omega_r}$, showing no evidence for an outlier across frequencies and loudspeaker positions.

Fig. 6 shows the relative computation time required to calculate $\phi(x)$ as a function of $\#F$. The reference (i.e., 100%) was the computation time for the high-resolution mesh. As expected, the computation time increased with $\#F$. On average across the tested $\#F$ s, the computation time was slightly larger for graded meshes than for uniform meshes. This might be a side-effect of our FMM implementation, which was optimized for uniform cluster sizes and thus uniform meshes. Compared to the UNI conditions, the POW1 condition showed the small-

	$\hat{\ell}$ (target)		ℓ (actual)			$\#\mathcal{F}$
	$\hat{\ell}_{min}$	$\hat{\ell}_{max}$	ℓ_{min}	ℓ_{max}	ℓ_{avg}	
UNI	2		1.1	2.8	2.0	77984
	3		1.8	4.4	2.9	35034
	5		3.0	6.6	4.8	12962
	7		3.9	9.9	7.0	5992
	10		6.2	13.4	10.0	2928
POW1	2	5	1.0	5.2	2.8	35162
	2	10	1.2	10.1	4.4	13286
	2	20	1.3	18.9	6.8	4820
POW4	2	20	1.0	18.7	2.5	32738
	2	40	1.0	35.9	2.6	23330
	3	40	1.5	31.9	3.8	12968
COS2	1	12	0.5	12.3	2.4	26416
	1	15	0.5	15.6	4.6	7494
	2	15	1.0	15.4	6.7	4730

Table 2: The SAP object: Mesh statistics for uniform (UNI) and graded (POW1, POW4, and COS2) meshes. Other details as in Tab. 1.

est increase in computation time, while the conditions POW2 and POW4 showed the largest increase (up to a factor of two) in the required computation time.

An interesting aspect is also the memory required for the calculations. The RAM consumption (for calculating $\phi(\mathbf{x})$ at 20 kHz) was 7.99 GB for the high-resolution mesh. For graded meshes with approximately 13 000 elements, the RAM consumption was 1.41 GB for the UNI mesh, 1.47 GB for the POW1 mesh, 1.52 GB for the POW2 mesh, and 1.74 GB for the COS2 mesh. Thus, mesh grading not only reduces the computation time, but also loosens the requirements on memory.

In summary, the three grading functions COS2, COS4, and POW4 showed the most potential in terms of smallest numerical errors in HRTF calculations. The POW1 grading showed the most potential in terms of smallest time required to calculate HRTFs.

4.5. The SAP object

The SAP object (SPH object with a generic pinna) was used to investigate the effect of an additional pinna on the mesh grading with a good relation to the results obtained for the SPH object. In order to create the SAP object, the default pinna from MakeHuman⁶ was stitched onto the left side, i.e., the positive y dimension, of the SPH object with the entrance of the ear canal at position $\mathbf{x} = [0, 100, 0]$ mm. The high-resolution mesh consisted of 140 847 elements with an AEL of approximately 1.4 mm. The high-resolution mesh was further re-meshed in order to obtain uniform meshes with AELs ranging from 2 to 10 mm. Also, the a-priori mesh grading was applied to the high-resolution mesh. The grading functions POW1, POS4, and COS2 were used. The COS4 and POW2 functions were

⁶MakeHuman: version 1.0.0, available from <http://www.makehuman.org> (date last viewed: January 31, 2016)

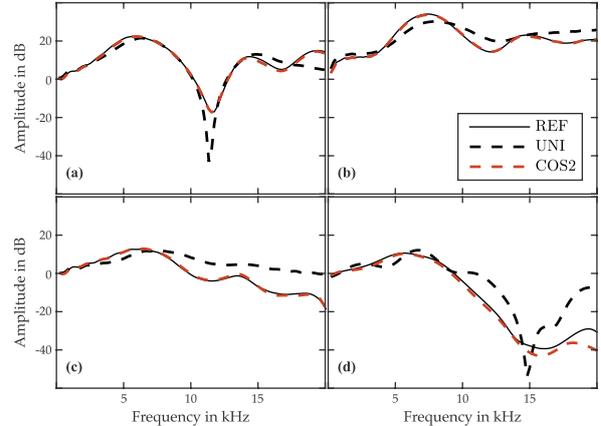


Figure 7: HRTF amplitude spectra calculated for the SPH object represented by the high-resolution mesh (REF, 108 940 elements), uniform mesh (UNI, 12 962 elements), and graded mesh (COS2, 7494 elements). The loudspeaker positions were (a) 0°, (b) 90°, (c) 180°, and (d) 270° (contralateral) in the horizontal plane.

not used because their previous results were very similar to those for COS2 and POW1, respectively. Table 1 shows the relevant mesh parameters for the tested grading functions and conditions.

For each mesh, a point source was placed at $\mathbf{x}^* = [0, 101, 0]$ mm and HRTFs were calculated using the EQA loud-speaker grid. The relative numerical errors with the high-resolution mesh HRTFs as reference were calculated.

Fig. 7 shows exemplary HRTFs of a uniform mesh (UNI, AEL of 5 mm) and a non-uniform mesh (COS2, minimum and maximum target edge length of 1 and 15 mm, respectively) calculated for a low $\#\mathcal{F}$ at four loudspeaker positions in the horizontal plane (front, left, rear, and right). The HRTFs calculated for the high-resolution mesh (REF) served as a reference in that comparison. While HRTFs of the uniform mesh differ in general at frequencies above 5 kHz, HRTFs of the graded mesh seem to correspond very well with the reference HRTFs. Only for the contralateral position, differences arise above 15 kHz but still these differences are in the range of 2 dB on average. For that position, HRTFs of the uniform mesh, show obvious differences, with an additional notch at 15 kHz. Note that in this comparison, a rather low $\#\mathcal{F}$ (about ten thousand elements) was used in order to 1) underline the limits of uniform meshes for the numerical HRTF calculation, and 2) evaluate the spectral details in HRTFs based on graded meshes.

Fig. 8 shows $e_{L^2}^{\Omega_c}$ calculated for all (a), ipsilateral (b), and contralateral (c) loudspeakers. Similar to results for the SPH object, the errors increased with decreasing $\#\mathcal{F}$. Also, for similar $\#\mathcal{F}$, POW1 and POW4 conditions yielded similarly smaller errors than the UNI condition. The COS2 condition yielded even smaller errors. The POW α conditions seem to require three times more and the UNI condition seems to require ten times more elements than the COS2 condition to achieve a similar relative error in the calculated HRTFs.

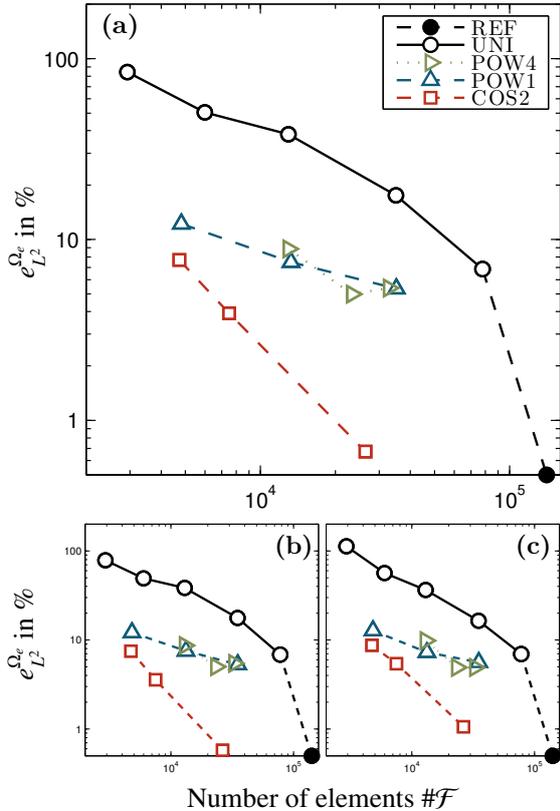


Figure 8: The SAP object: Relative numerical errors for (a) all nodes, (b) ipsilateral nodes, and (c) contralateral nodes of the EQA grid. The reference for the errors was the HRTF set calculated for the high-resolution mesh.

Averaged across all conditions, the $e_{L^2}^{\Omega_c}$ was 1.09 times larger than $e_{L^2}^{\Omega_c}$ showing a very small variance and no evidence for an outlier across the loudspeaker positions.

In summary, the meshes graded with the POW4 function yielded errors similar to those of the meshes graded with the POW1 function. The COS2 grading function showed the most promising results. For this function, the HRTF spectra did not show any problematic issues. And even for the coarsest COS2 graded mesh (4730 elements), the error was smaller than that of HRTFs calculated for the finest uniform mesh (77 984 elements, AEL of 2 mm). This is interesting because the minimum edge length of that COS2 graded mesh was similar to the edge length of that uniform mesh (see ℓ_{min} in Table 2), implying that while both meshes had a similar resolution at the ipsilateral side, the graded mesh had a much coarser resolution at the contralateral side. This supports again our assumption that a fine mesh resolution at the ipsilateral pinna is important, whereas for the rest of the mesh, a coarser resolution can still yield acceptable accuracy.

4.6. The HUM objects

The HUM objects were geometric models of the head and pinnae of three actual listeners (NH5, NH130, and NH131).

	$\hat{\ell}$ (target)		ℓ (actual)			$\#\mathcal{F}$
	$\hat{\ell}_{min}$	$\hat{\ell}_{max}$	ℓ_{min}	ℓ_{max}	ℓ_{avg}	
UNI	2		0.3	3.8	1.9	98009
	3		1.5	4.3	2.8	50115
	4		2.2	5.6	3.7	28320
	5		2.8	19.7	5.2	16630
	10		2.9	19.7	6.3	4689
POW1	2	5	1.0	5.5	3.0	39447
	2	8	1.1	8.4	3.6	25774
	2	10	1.1	10.5	5.2	12407
	2	12	1.1	12.4	5.7	9768
COS2	1	5	0.5	5.2	2.3	57526
	1	8	0.5	8.4	2.9	30697
	1	15	0.5	16.5	3.7	13722

Table 3: The HUM objects: Mesh statistics for uniform (UNI) and graded (POW1 and COS2) meshes averaged across subjects and ears. Other details as in Tab. 1.

These high-resolution meshes originate from scanning the head with a laser scanner and scanning silicone impressions of listener's pinnae in a high-energy industrial computer-tomography scanner (Reichinger et al., 2013). The meshes consisted of 111 362, 111 422 and 107 692 elements for NH5, NH130, and NH131, respectively, and are described in more details in Ziegelwanger et al. (2015b).

For each listener, the high-resolution mesh was re-meshed in order to obtain uniform meshes with AELs ranging from 2 mm to 10 mm. Graded meshes were created by applying the a-priori mesh grading to the high-resolution meshes⁷. Based on the results from Sec. 4.4 and 4.5, the grading functions POW1 and COS2 were considered only. Table 3 shows relevant mesh parameters for the tested grading functions and conditions. For each mesh, HRTFs were calculated for both grids (ARI and EQA). Then, the relative numerical errors with the high-resolution mesh HRTFs as reference were calculated.

Fig. 9 shows HRTFs of NH5 calculated for the meshes with between 10 000 and 20 000 elements and four loudspeaker positions as in Fig. 7. The high-resolution mesh HRTFs are shown as a reference. HRTFs calculated for the COS2-graded meshes seem to correspond very well to HRTFs of the high-resolution mesh, except for minor differences at the contralateral loudspeaker position above 15 kHz. HRTFs calculated for the UNI mesh show much more deviation, e.g., an additional notch at around 10 kHz arose at the ipsilateral and rear loudspeaker positions.

Fig. 10 shows $e_{L^2}^{\Omega_c}$ averaged across subjects and ears and calculated for all (a), ipsilateral (b), and contralateral (c) loudspeakers of the EQA grid. For uniform meshes, the errors increased with decreasing $\#\mathcal{F}$, reaching a maximum at around 100% for $\#\mathcal{F}$ of approximately 28 000, and saturating at this level for less elements. For the POW1-graded meshes, the er-

⁷Note that for each ear separate meshes were created, yielding two meshes per subject and condition.

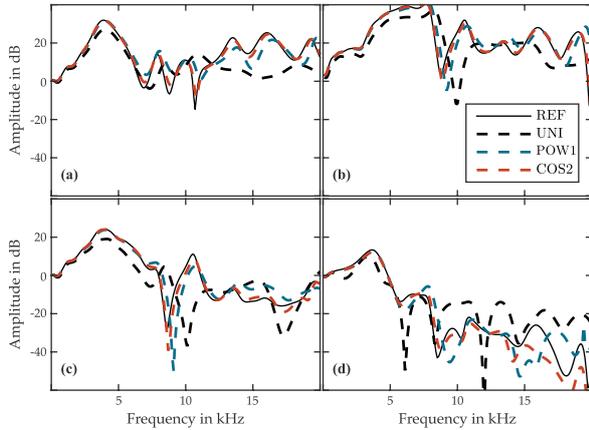


Figure 9: HRTF amplitude spectra calculated for the HUM object of NHS from the ARI database, represented by the high-resolution mesh (REF, 111 362 elements), a uniform mesh (UNI, 17 023 elements), a linearly graded mesh (POW1, 12 041 elements), and raised cosine graded mesh (COS2, 13 633 elements). The loudspeaker positions were (a) 0° , (b) 90° , (c) 180° , and (d) 270° in the horizontal plane.

rors increased with decreasing $\#\mathcal{F}$, but the effect of $\#\mathcal{F}$ was not that large as for uniform meshes and for other objects. For the COS2-graded meshes, the effect of $\#\mathcal{F}$ was more pronounced. Compared to the errors in the UNI condition, the relative error in the COS2 condition was approximately ten times smaller. For the coarsest COS2-graded mesh ($\#\mathcal{F} \approx 13\,000$), the error was even slightly smaller than for the finest uniform mesh ($\#\mathcal{F} \approx 98\,000$).

For the psychoacoustic-motivated evaluation, acoustically measured HRTFs of the corresponding listeners from the ARI database were used as references⁸. These HRTFs were measured in a semi-anechoic chamber by placing in-ear microphones in the listeners' blocked ear canals and applying the multiple exponential sweep method (Majdak et al., 2007) for the system identification of each HRTF direction. These HRTFs were available at positions described by the ARI grid (for more details on the measurement see Majdak et al., 2010; Ziegelwanger et al., 2015b).

First, broadband temporal features were evaluated by means of equivalent head radius derived for an HRTF set (for each ear separately) from the TOA model. Table 4 shows the equivalent head radii obtained for all conditions, listeners, and ears. For the most of the numerically calculated HRTFs, the equivalent head radii were ± 2 mm around those of measured HRTFs, showing no evidence for artifacts larger than a few μ s in the broadband timing of the calculated HRTFs. The equivalent head radii also did not change much when $\#\mathcal{F}$ was reduced, indicating that the proposed mesh grading did not introduce critical artifacts in the broadband temporal features of calculated HRTFs.

Second, spectral features were evaluated by means of sound-

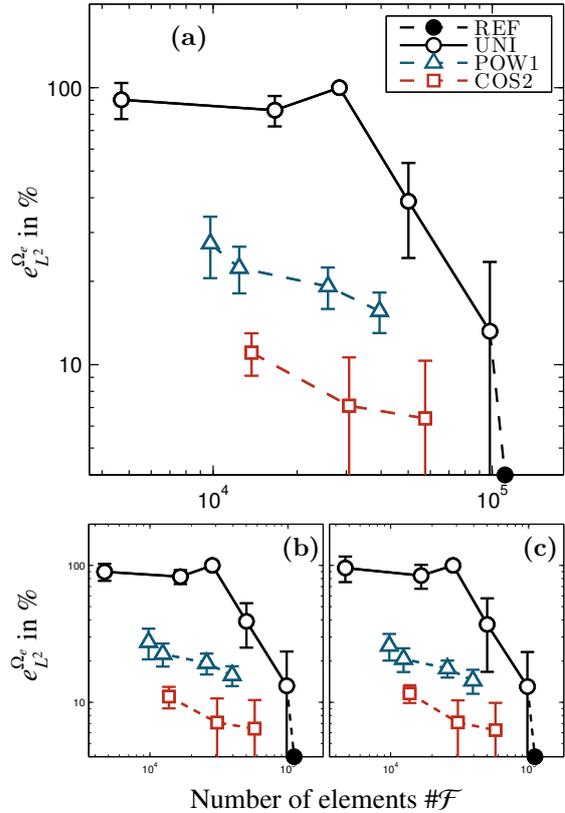


Figure 10: HUM objects: Relative numerical errors averaged across the listeners and ears for (a) all nodes, (b) ipsilateral, and (c) contralateral positions of the EQA grid. Reference for the error was the HRTF set calculated for the high-resolution mesh of each individual listener and ear.

localization performance predicted for the calculated HRTFs by the sagittal-plane sound-localization model. The performance predictions were calculated for the three listeners. Besides the reference HRTFs (which were the acoustically measured HRTFs), the model requires a parameter called 'uncertainty', representing the ability of a listener to localize sounds (Baumgartner et al., 2013). We used an uncertainty of 1.9, corresponding to an average localizer. Four benchmarks were used for the analysis. The first benchmark, 'ACTUAL', was the actual localization performance of 14 human listeners (Middlebrooks, 1999). Comparison to that benchmark allows to estimate the quality of predictions relative to the actual localization performance. The second benchmark, 'OWN', was the localization performance predicted for 177 listeners from three HRTF databases (ARI, LISTEN, CIPIC)⁹ localizing sound sources with their own ears' HRTFs. Comparison to that benchmark allows to estimate the usual range of predictions when localizing with own ears. The third benchmark, 'OTHER' was the localization performance predicted for our three listeners

⁸Available from <http://www.kfs.oew.ac.at/hrtf> (date last viewed: January 31, 2016)

⁹Available from <http://sofaconventions.org> (date last viewed: January 31, 2016)

		Equivalent head radius in mm						
$\hat{\ell}$ (target)		NH5		NH130		NH131		
$\hat{\ell}_{min}$	$\hat{\ell}_{max}$	l	r	l	r	l	r	
AC	-	91.8	92.4	92.4	92.0	94.5	94.8	
	1	90.8	90.8	89.2	89.5	94.6	94.6	
UNI	2	91.4	91.4	89.4	89.7	94.6	94.6	
	3	91.4	91.7	89.4	89.4	94.6	94.6	
	4	91.8	91.6	90.2	90.2	94.5	94.5	
	5	91.1	91.1	89.8	89.8	95.1	95.1	
	10	88.2	89.0	88.3	88.5	92.8	92.8	
POW1	2	91.3	91.2	89.3	89.3	94.7	94.7	
	2	8	90.9	90.9	89.5	89.5	94.7	94.7
	2	10	91.4	91.3	89.4	89.4	94.3	94.3
	2	12	90.9	90.7	88.8	89.1	94.6	94.6
COS2	1	5	90.7	90.7	89.1	89.1	94.6	94.6
	1	8	90.7	90.6	89.0	89.2	94.3	94.3
	1	15	90.3	90.3	88.8	88.8	94.2	94.2

Table 4: Equivalent head radii calculated separately for the (l)eft and (r)ight ear of NH5, NH130, and NH131 from the ARI database. Acoustically measured HRTFs (AC) and HRTFs numerically calculated for uniform (UNI) and graded meshes (LIN and COS2). Other details as in Tab. 1.

localizing sound sources with 176 others' HRTF sets from the databases. This benchmark represents the result of localizing with non-individual HRTFs. Reaching this error level indicates no need for individual HRTFs in the corresponding condition (for more details see Ziegelwanger et al., 2015b). The fourth benchmark, 'REF' was the localization performance predicted for the three listeners localizing sources with the HRTFs calculated for the high-resolution mesh. Comparison to this benchmark allows to estimate the ultimate effect of re-meshing on the sound-localization performance.

Fig. 11 shows the predicted localization performance as functions of $\#\mathcal{F}$ along with the benchmarks. The benchmarks 'OWN' and 'OTHER' were within and outside the actual listener performance represented by the benchmark 'ACTUAL', respectively, implying a good level of validity for the predictions. For each of the listeners, the benchmark 'REF' was within the range of 'ACTUAL', confirming the general ability of numerically calculated HRTFs to replace acoustically measured HRTFs. For uniform meshes, the performance started at the level of the benchmark 'REF' for the finest mesh and degraded with decreasing $\#\mathcal{F}$. For the coarsest uniform mesh, the predictions were at (or beyond) the level of the benchmark 'OTHER', indicating that these individual HRTFs did not provide any advantages over non-individual HRTFs. For graded meshes, the performance also started at the level of the benchmark 'REF' and degraded with decreasing $\#\mathcal{F}$. In contrast to the uniform meshes, the effect of $\#\mathcal{F}$ was much smaller and the performance was still within the range of benchmark 'OWN' down to meshes with $\#\mathcal{F}$ of approximately 13 000. Compared to uniform meshes, the mesh grading allowed to reduce $\#\mathcal{F}$ by a factor of approximately seven without introducing a degrada-

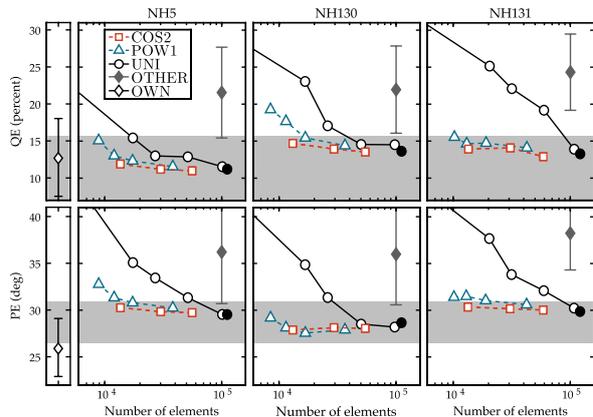


Figure 11: Effect of re-meshing on the sagittal-plane localization performance by means of quadrant error (QE) and polar error (PE). Most-left column: Benchmark 'OWN'. Other columns (NH5, NH130, and NH131): Predictions for the corresponding listener localizing with own-ear HRTFs calculated for the uniform (open circles) and for the non-uniform (triangles) meshes as functions of $\#\mathcal{F}$. Filled circles: Benchmark 'REF'. Filled diamonds: Benchmark 'OTHER'. Grey area: benchmark 'ACTUAL'. See text for details on the benchmarks.

tion of the predicted performance. When compared across the two grading functions, the COS2 grading seems to provide a slightly better performance than the POW1 grading.

In order to roughly quantify the relation between the predicted localization performance and $\#\mathcal{F}$, linear regressions were fit to the predictions. Similar to the presentation of the abscissa in Fig. 11, the fits were performed on $\log_{10}(\#\mathcal{F})$ yielding a PE slope (in $^{\circ}/\log\#\mathcal{F}$) and a QE slope (in $\%/ \log\#\mathcal{F}$) per listener and mesh condition. The individual slopes and their averages are shown in Table 5. For example, for NH5 and uniform meshes, the PE slope of $-11.2^{\circ}/\log\#\mathcal{F}$ indicates that by increasing $\#\mathcal{F}$ by the factor of ten, the PE decreased by 11.2° .

Our results show that when reducing $\#\mathcal{F}$ in the uniform mesh by factor of ten, the PE and QE errors increased by 10° and 10% , respectively. In contrast, for the COS2-graded meshes, the same reduction of elements increased these errors by 0.23° and 1.7% , respectively, only.

		PE ($^{\circ}/\log\#\mathcal{F}$) and QE ($\%/ \log\#\mathcal{F}$) slopes							
		NH5		NH130		NH131		Avg.	
		PE	QE	PE	QE	PE	QE	PE	QE
UNI		-11.2	-8.2	-9.8	-11.0	-10.0	-13.2	-10.2	-10.5
POW1		-3.6	-4.9	-1.7	-7.5	-1.4	-1.9	-1.8	-5.0
COS2		-0.9	-1.5	0.3	-1.9	-0.5	-1.5	-0.2	-1.7

Table 5: Slopes of the predicted sound-localization performance parameters (PE and QE) for NH5, NH130, and NH131 from the ARI database and the average across subjects. HRTFs numerically calculated for uniform (UNI) and graded meshes (POW1 and COS2). Other details as in Tab. 1.

5. Conclusions

A mesh preprocessing method for the numerical calculation of HRTFs, i.e., *a-priori mesh grading*, was proposed. For the evaluation, the method was applied to meshes of three geometric objects with various grading functions. HRTFs were calculated based on these meshes and the results were evaluated by means of numerical errors and perceptually motivated model predictions.

When comparing the HRTFs calculated for graded meshes with those calculated for uniform meshes, the HRTF calculation for graded meshes yielded similar or even better results than for high-resolution uniform meshes in terms of numerical accuracy and in terms of the predicted sound-localization performance. Thus, graded meshes seem to be suitable for the numerical calculation of HRTFs in the full audible frequency range, further indicating that the recommendation of at least six elements per wavelength can be violated apart from the microphone area and the ipsilateral pinna.

The numerical accuracy of HRTFs calculated for various grading functions was compared including linear grading, higher-order power grading, and raised-cosine based grading. For the simple geometric object, the sphere, all grading functions showed better performance than uniform re-meshing in terms of the relative numerical error and computation time. The raised-cosine based grading and fourth-order grading functions showed the most potential. HRTFs calculated for these grading functions showed smaller errors than HRTFs calculated for the high-resolution mesh - even with less elements in the meshes. These grading functions concentrate more elements in the proximity range of the microphone area and as a consequence, less elements in total are required to obtain similar numerical errors in the numerical calculation process. Raised-cosine based grading showed the best overall performance. With approximately 13 000 elements, the computation took only ten percent of the computation required for a high-resolution mesh, but still the relative numerical and the perceptual errors were in the range of that for the high-resolution mesh.

The proposed *a-priori* mesh grading algorithm was implemented as a plug-in for OpenFlipper and can be combined with the Mesh2HRTF package to calculate HRTFs. In that implementation, the proposed mesh grading algorithm reduced the calculation time and memory requirements significantly. These computational costs might be further reduced by optimizing the FMM for graded meshes and introducing a frequency-specific mesh grading in which the grading parameters are further optimized for the simulated frequency.

Acknowledgements

This study was supported by the Austrian Science Fund (FWF, projects P 24124 and I 1018 N25). The authors want to thank Andreas Reichinger (VRVis GmbH, Austria) and his team for the work on geometry acquisition required to obtain the high-resolution meshes of the HUM objects.

References

- Baumgartner, R., Majdak, P., Laback, B., 2013. Assessment of Sagittal-Plane Sound Localization Performance in Spatial-Audio Applications. In: *The Technology of Binaural Listening*. Springer, Berlin, DE, pp. 93–119.
URL http://dx.doi.org/10.1007/978-3-642-37762-4_4
- Beranek, L. L., Mellow, T. J., Aug. 2012. *Acoustics: Sound Fields and Transducers*. Academic Press, Amsterdam, Netherlands.
- Botsch, M., Kobbelt, L., 2004. A Remeshing Approach to Multiresolution Modeling. In: *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing. SGP '04*. ACM, New York, NY, USA, pp. 185–192.
URL <http://dx.doi.org/10.1145/1057432.1057457>
- Burton, A. J., Miller, G. F., Jun. 1971. The Application of Integral Equation Methods to the Numerical Solution of Some Exterior Boundary-Value Problems. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 323 (1553), 201–210.
URL <http://dx.doi.org/10.1098/rspa.1971.0097>
- Chen, J. T., Chen, K. H., Chen, C. T., May 2002. Adaptive boundary element method of time-harmonic exterior acoustics in two dimensions. *Computer Methods in Applied Mechanics and Engineering* 191 (31), 3331–3345.
URL [http://dx.doi.org/10.1016/S0045-7825\(02\)00214-1](http://dx.doi.org/10.1016/S0045-7825(02)00214-1)
- Chen, Z.-S., Waubke, H., Kreuzer, W., Jun. 2008. A formulation of the fast multipole boundary element method (FMBEM) for acoustic radiation and scattering from three-dimensional structures. *Journal of Computational Acoustics* 16 (2), 303–320.
URL <http://dx.doi.org/10.1142/S0218396X08003725>
- Gaul, L., Kögl, M., Wagner, M., Feb. 2003. *Boundary Element Methods for Engineers and Scientists: An Introductory Course with Advanced Topics*. Springer, Berlin, DE.
URL <http://dx.doi.org/10.1007/978-3-662-05136-8>
- Goldstein, C. I., Feb. 1982. The finite element method with non-uniform mesh sizes applied to the exterior Helmholtz problem. *Numerische Mathematik* 38 (1), 61–82.
URL <http://dx.doi.org/10.1007/BF01395809>
- Gumerov, N. A., O'Donovan, A. E., Duraiswami, R., Zotkin, D. N., Jan. 2010. Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. *The Journal of the Acoustical Society of America* 127 (1), 370–386.
URL <http://dx.doi.org/10.1121/1.3257598>
- Heinrich, W., Beilenhoff, K., Mezzanotte, P., Roselli, L., Sep. 1996. Optimum mesh grading for finite-difference method. *IEEE Transactions on Microwave Theory and Techniques* 44 (9), 1569–1574.
URL <http://dx.doi.org/10.1109/22.536606>
- Hunter, P., Pullan, A., 2002. *FEM/BEM Notes*.
- Lin, C. T., Guillon, P., Epain, N., Zolfaghari, R., van Schaik, A., Tew, A. I., Hetherington, C., Thorpe, J., Jan. 2014. Creating the Sydney York Morphological and Acoustic Recordings of Ears Database. *IEEE Transactions on Multimedia* 16 (1), 37–46.
URL <http://dx.doi.org/10.1109/TMM.2013.2282134>
- Kahana, Y., Nelson, P. A., Apr. 2006. Numerical modelling of the spatial acoustic response of the human pinna. *Journal of Sound and Vibration* 292 (1–2), 148–178.
URL <http://dx.doi.org/10.1016/j.jsv.2005.07.048>
- Kahana, Y., Nelson, P. A., Mar. 2007. Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models. *Journal of Sound and Vibration* 300 (3–5), 552–579.
URL <http://dx.doi.org/10.1016/j.jsv.2006.06.079>
- Katz, B. F. G., Nov. 2001a. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *The Journal of the Acoustical Society of America* 110 (5), 2440–2448.
URL <http://dx.doi.org/10.1121/1.1412440>
- Katz, B. F. G., Nov. 2001b. Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements. *The Journal of the Acoustical Society of America* 110 (5), 2449–2455.
URL <http://dx.doi.org/10.1121/1.1412441>
- Kreuzer, W., Majdak, P., Chen, Z., Sep. 2009. Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range. *The Journal of the Acoustical Society of America* 126 (3), 1280–1290.
URL <http://dx.doi.org/10.1121/1.3177264>

- Langer, U., Mantzafaris, A., Moore, S. E., Touloupoulos, I., 2015. Mesh grading in isogeometric analysis. *Computers & Mathematics with Applications* 70 (7), 1685–1700.
URL <http://dx.doi.org/10.1016/j.camwa.2015.03.011>
- Liang, M. T., Chen, J. T., Yang, S. S., Mar. 1999. Error estimation for boundary element method. *Engineering Analysis with Boundary Elements* 23 (3), 257–265.
URL [http://dx.doi.org/10.1016/S0955-7997\(98\)00086-1](http://dx.doi.org/10.1016/S0955-7997(98)00086-1)
- Luebke, D., May 2001. A developer's survey of polygonal simplification algorithms. *IEEE Computer Graphics and Applications* 21 (3), 24–35.
URL <http://dx.doi.org/10.1109/38.920624>
- Macpherson, E. A., Middlebrooks, J. C., May 2002. Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *The Journal of the Acoustical Society of America* 111 (5), 2219–2236.
URL <http://dx.doi.org/10.1121/1.1471898>
- Majdak, P., Balazs, P., Laback, B., Jul. 2007. Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions. *Journal of the Audio Engineering Society* 55 (7/8), 623–637.
URL <http://www.aes.org/e-lib/browse.cfm?elib=14190>
- Majdak, P., Goupell, M. J., Laback, B., Feb. 2010. 3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training. *Attention, perception & psychophysics* 72 (2), 454–469.
URL <http://dx.doi.org/10.3758/APP.72.2.454>
- Marburg, S., Mar. 2002. Six boundary elements per wavelength: Is that enough? *Journal of Computational Acoustics* 10 (1), 25–51.
URL <http://dx.doi.org/10.1142/S0218396X02001401>
- Mehrgardt, S., Mellert, V., Jun. 1977. Transformation characteristics of the external human ear. *The Journal of the Acoustical Society of America* 61 (6), 1567–1576.
URL <http://dx.doi.org/10.1121/1.381470>
- Middlebrooks, J. C., Sep. 1999. Individual differences in external-ear transfer functions reduced by scaling in frequency. *The Journal of the Acoustical Society of America* 106 (3), 1480–1492.
URL <http://dx.doi.org/10.1121/1.427176>
- Morimoto, M., Aokata, H., 1984. Localization cues of sound sources in the upper hemisphere. *The Journal of the Acoustical Society of Japan* (5), 166–173.
URL <http://dx.doi.org/10.1250/ast.5.165>
- Morse, P. M., Ingard, K. U., 1986. *Theoretical acoustics*. Princeton University Press, Princeton, NJ.
- Möbius, J., Kobbelt, L., 2012. OpenFlipper: An Open Source Geometry Processing and Rendering Framework. In: *Proceedings of the 7th International Conference on Curves and Surfaces*. Springer-Verlag, Berlin, DE, pp. 488–500.
URL http://dx.doi.org/10.1007/978-3-642-27413-8_31
- Møller, H., 1992. Fundamentals of binaural technology. *Applied Acoustics* 36 (3–4), 171–218.
URL [http://dx.doi.org/10.1016/0003-682X\(92\)90046-U](http://dx.doi.org/10.1016/0003-682X(92)90046-U)
- Møller, H., Sørensen, M. F., Hammershøi, D., Jensen, C. B., May 1995. Head-Related Transfer Functions of Human Subjects. *Journal of the Audio Engineering Society* 43 (5), 300–321.
URL <http://www.aes.org/e-lib/browse.cfm?elib=7949>
- Reichinger, A., Majdak, P., Sablatnig, R., Maierhofer, S., Jun. 2013. Evaluation of Methods for Optical 3-D Scanning of Human Pinnae. In: *2013 International Conference on 3D Vision - 3DV 2013*. pp. 390–397.
URL <http://dx.doi.org/10.1109/3DV.2013.58>
- Rui, Y., Yu, G., Xie, B., Liu, Y., May 2013. Calculation of Individualized Near-Field Head-Related Transfer Function Database Using Boundary Element Method. In: *Proceedings of the 134th Convention of the Audio Engineering Society*. Rome, IT.
- Søndergaard, P., Majdak, P., 2013. The Auditory Modeling Toolbox. In: *Blauert, J. (Ed.), The Technology of Binaural Listening*. Springer, Berlin, DE, pp. 33–56.
URL http://dx.doi.org/10.1007/978-3-642-37762-4_2
- Treeby, B. E., Pan, J., Nov. 2009. A practical examination of the errors arising in the direct collocation boundary element method for acoustic scattering. *Engineering Analysis with Boundary Elements* 33 (11), 1302–1315.
URL <http://dx.doi.org/10.1016/j.enganabound.2009.06.005>
- Walsh, T., Demkowicz, L., Jan. 2003. hp Boundary element modeling of the external human auditory system—goal-oriented adaptivity with multiple load vectors. *Computer Methods in Applied Mechanics and Engineering* 192 (1–2), 125–146.
URL [http://dx.doi.org/10.1016/S0045-7825\(02\)00536-4](http://dx.doi.org/10.1016/S0045-7825(02)00536-4)
- Wightman, F. L., Kistler, D. J., Feb. 1989. Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America* 85 (2), 858–867.
URL <http://dx.doi.org/10.1121/1.397557>
- Ziegelwanger, H., Majdak, P., Mar. 2014. Modeling the direction-continuous time-of-arrival in head-related transfer functions. *The Journal of the Acoustical Society of America* 135 (3), 1278–1293.
URL <http://dx.doi.org/10.1121/1.4863196>
- Ziegelwanger, H., Majdak, P., Kreuzer, W., 2015a. Mesh2hrtf: Open-source software package for the numerical calculation of head-related transfer functions. In: *Proceedings of the 22nd International Congress of Sound and Vibration*. Florence, IT.
- Ziegelwanger, H., Majdak, P., Kreuzer, W., Jul. 2015b. Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization. *The Journal of the Acoustical Society of America* 138 (1), 208–222.
URL <http://dx.doi.org/10.1121/1.4922518>

Chapter 8

Efficient numerical calculation of HRTFs

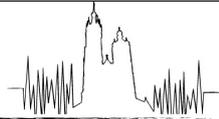
This work was published as:

Ziegelwanger, H., Majdak, P., and Kreuzer, W. (2014). "Efficient numerical calculation of head-related transfer functions," in *Proceedings of the Forum Acusticum 2014*, Krakow, PL, 1-6. [doi:10.13140/RG.2.1.2165.8645](https://doi.org/10.13140/RG.2.1.2165.8645)

Reducing the computation time in the numerical HRTF calculation process by near-field calculation and range extrapolation of HRTFs was the idea of me, the first author, and the third author. The design of the study benefited from fruitful discussions with the second author. I implemented the range extrapolation, numerically calculated the HRTFs, evaluated the results and wrote the initial draft of the manuscript. The third author wrote section 2.5 (spherical-harmonics representation and range extrapolation of HRTFs). The co-authors helped writing the manuscript.

Co-author 1:

Co-author 2:



Efficient Numerical Calculation of Head-Related Transfer Functions

Harald Ziegelwanger, Piotr Majdak, Wolfgang Kreuzer
Acoustics Research Institute, Austrian Academy of Sciences, Vienna, Austria.

Summary

The directional filtering of the incoming sound at the ear canal can be described by the head-related transfer functions (HRTFs). HRTFs are caused by reflection, scattering and shadowing effects of the particular geometry of a listener's head, torso and pinna. While listener-specific HRTFs are usually measured acoustically, they can also be numerically calculated based on listener's geometry. The general recommendation for the discretization of a listener's geometry, i.e., 3-D mesh, is to use at least 6 boundary elements per wavelength in order to keep the numerical error small and the calculation stable. The numeric formulation results in a densely populated linear system of equations with up to 150 000 unknowns and the calculation of HRTFs lasts tens of hours. In this study, methods for reducing the numerical costs are proposed. First, a non-uniform geometry re-meshing algorithm reducing the size of the equation system is proposed. Second, calculation of near-field HRTFs and range-extrapolated to far-field HRTFs is proposed. The methods were evaluated in terms of the modeled sound-localization performance. Our results show a clear advantage of a non-uniform geometry discretization in the process of HRTF simulation.

PACS no. 43.66.Qp, 43.66.Pn, 43.20.Rz, 43.20.Fn

1. Introduction

The directional filtering of the incoming sound at the entrance of the ear-canal is described by the head-related transfer functions (HRTFs) [1]. HRTFs are usually measured acoustically [2], but they can also be calculated numerically with the boundary element method (BEM) [3]. The BEM solves the boundary integral equation (BIE), i.e., Helmholtz equation for a boundary surface (a listener's geometry), in the three-dimensional (3D) space numerically.

Practically, our collocation-BEM-implementation [4] consists of three processing stages. First, a linear system of equations is set up (Sec. 2.4.2) based on the BIE for the discretized boundary surface, i.e., a 3D polygon mesh. It is generally recommended to discretize the geometry uniformly and with at least six elements per wavelength [5] in order to keep the numerical error small and the simulation process stable. For human listeners, the pinna requires a discretization of an average edge length (AEL) of 1 mm in order to obtain perceptively valid HRTFs in terms of adequate sound localization performance [6]. This criterion for simulations in the entire audible frequency range results in up to 150 000 degrees of freedom

(DOF). Second, the sound-field on the surface is calculated by solving the equation system (Sec. 2.4.3). Third, the sound-field is evaluated for the positions of sound sources (Sec. 2.4.4) yielding in an HRTF set.

The computational effort for all three stages is determined by the DOF and can be reduced by techniques like the principle of reciprocity and the multi-level fast-multipole method (ML-FMM) [3, 7]. Nevertheless, the simulation process might still last tens of hours. For the two-dimensional Helmholtz equation, the DOF can be reduced by a non-uniform discretization without affecting the stability of the simulation and without increasing the simulation error [8]. Recently, for 3D polygon meshes, a non-uniform discretization algorithm based on [8] was proposed [9]. The algorithm aims at reduction of the element number in certain regions of the mesh and its applicability was already shown for HRTF simulations [10].

In this study, we further evaluate the effect of the non-uniform discretization on sound localization for HRTFs of 3 human listeners, for whom HRTFs were acoustically measured and numerically simulated. For the simulations, uniform and non-uniform meshes were obtained for those listeners, with 9 meshes per listener. Further, since the third BEM stage additionally depends on the source-ear distance, we propose to improve the simulation efficiency by simulating near-field HRTFs and range-extrapolating to far-field HRTFs using spherical-harmonics [11, 7].

2. Methods

2.1. Acoustic Measurement

Acoustically measured HRTFs of three listeners (NH5, NH130 and NH131 of the ARI-HRTF-database¹) were used. In a semi-anechoic chamber, in-ear microphones were placed in the listeners' ear canals and multiple exponential sweep method [2] was used to measure HRTFs for 1550 directions in a distance of 1.2 m. The measured transfer functions were equalized and directional transfer functions (DTFs) were calculated in order to emphasize the direction-dependent information in HRTFs (for more details see [12]).

2.2. Geometry Acquisition

Two geometry acquisition methods were used to capture the geometry of the head and the pinnae. First, the head was scanned by a laser scanner and the mesh was extracted from an internal volumetric representation. The neck was cut and closed. Second, a high-energy industrial computer-tomography scanner was used to capture the geometry of silicone impressions of the pinnae. The pinna meshes were merged with the head mesh by manually aligning the pinnae in the meshes. The seam region was re-sculpted and smoothed. The final high-accuracy mesh was watertightened and consisted of approximately 2.5 million triangles (for more details see [6]).

2.3. Re-meshing

A reference mesh was created by re-meshing the high-accuracy mesh to a mesh with an AEL of 1 mm at the pinna and 2.5 mm at the rest of the head surface (Fig. 2 a, b). Note that for the left and right ear HRTFs, two different meshes were used. For the simulations, the contra-lateral ear was cut out and the resulting hole was filled.

2.3.1. Uniform Meshes

For each listener the high-accuracy mesh was re-meshed to uniformly discretized meshes with AELs of 2 to 5 mm (Fig. 2 c, d) by applying a uniform re-meshing algorithm [13] implemented in [14].

2.3.2. Non-Uniform Meshes

For the non-uniform mesh, we loose the recommendation of at least 6 elements per wavelength by assuming that larger element size in regions far apart from the modeled sound source, i.e., at the contra-lateral side, has only an irrelevant impact on the simulated HRTF. This is justified as followed. First, the Green's function and its derivatives decay fast with increasing distance and, thus, errors arising from boundary

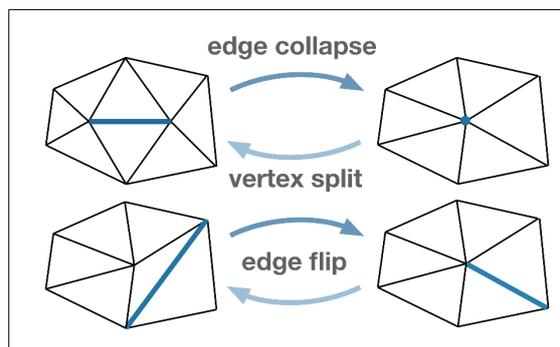


Figure 1. Illustration of edge collapse, vertex split and edge flip

elements at the contra-lateral side barely affect the results of boundary elements near to the modeled sound source. Second, any error caused by the increased element size at the contra-lateral side is damped by a factor in the order of tens of dB, having only a small effect on the final HRTF (for more details see [10]).

Based on the implementation of the uniform re-meshing algorithm [13], a non-uniform re-meshing algorithm was created. The algorithm was implemented as a module in Openflipper [14]. The module requires two input parameters: 1) the preferred minimum edge length at the entrance of the ear canal and 2) the maximal edge length at the contra-lateral side. The algorithm, then, increases the edge length as a linear function of the distance of an edge to the entrance of the ear canal in an iterative procedure. Within one iteration, 1) short edges are collapsed to points (Fig. 1) and, thus, neighboring edges are stretched, 2) long edges are shortened by splitting an endpoint (Fig. 1), 3) the longest edges of obtuse-angled triangles are flipped (Fig. 1) in order to regularize the mesh, and 4) the mesh is smoothed and nodes are projected onto the original, uniform mesh.

In this study, that algorithm was applied on the high-accuracy mesh of each listener using the parameters listed in Tab. I. Usually, ten iterations were sufficient to achieve the required edge length distribution (Fig. 2 e, f).

2.4. Numeric Simulation

2.4.1. Boundary integral equation

Assuming a time factor of $e^{-i\omega t}$ the main equation for acoustics in an exterior domain Ω_e outside an object Ω with boundary Γ is the Helmholtz equation:

$$\begin{aligned} \nabla^2 \phi(\mathbf{x}) + k^2 \phi(\mathbf{x}) &= 0, & \mathbf{x} \in \Omega_e \\ \frac{\partial \phi(\mathbf{x})}{\partial \mathbf{n}} &= v(\mathbf{x}), & \mathbf{x} \in \Gamma \end{aligned}$$

$$\lim_{|\mathbf{x}| \rightarrow \infty} |\mathbf{x}| \left(\frac{\partial \phi(\mathbf{x})}{\partial |\mathbf{x}|} - ik\phi(\mathbf{x}) \right) = 0 \quad (1)$$

¹ Available from <http://www.kfs.oew.ac.at/hrtf> (date last viewed May 27, 2014)

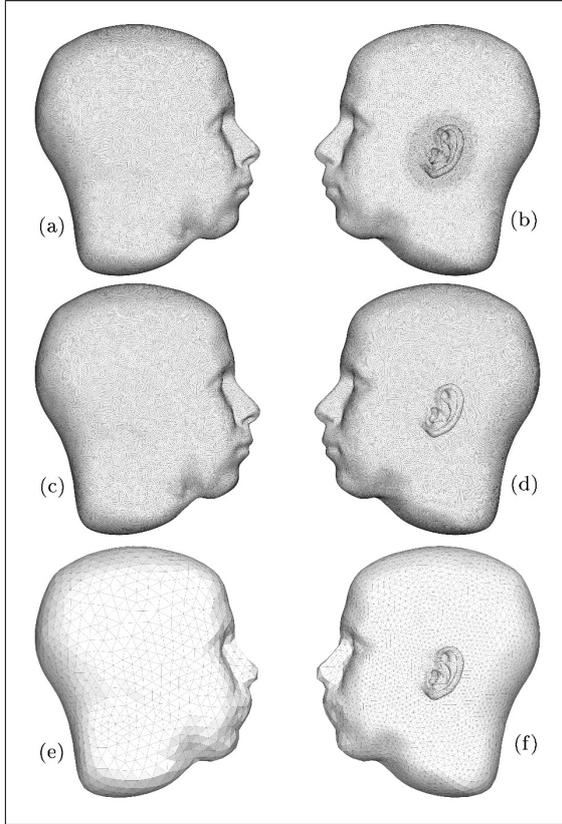


Figure 2. Exemplary meshes (left ear, NH5). Contra-lateral (a, c, e) and ipsi-lateral (b, d, f) view. Reference (a, b), uniform (c, d), and non-uniform (e, f) mesh.

where $\phi(\mathbf{x}) = \frac{p(\mathbf{x})}{i\omega\rho}$ denotes the velocity potential at the point \mathbf{x} and $k = \frac{\omega}{c}$ is the wavenumber dependent on the speed of sound c and the circular frequency ω . $p(\mathbf{x})$ represents the (complex) sound pressure at a point \mathbf{x} , ρ the density of the medium and i the imaginary unit. Γ is the boundary surface (here, head and ears), \mathbf{n} the vector normal to Γ at a point \mathbf{x} pointing to Ω_e and $v(\mathbf{x}) = \nabla v(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})$ the particle velocity in direction of \mathbf{n} . Eq. (1) is transformed into a boundary integral equation (BIE). By applying the Burton-Miller method [15], the BIE for the exterior Helmholtz problem reads as:

$$-\frac{1}{2}\phi(\mathbf{x}) + L[\phi](\mathbf{x}) + \frac{i}{k} \frac{\partial}{\partial \mathbf{n}_{\mathbf{x}}} L[\phi](\mathbf{x}) = \frac{i}{k} \frac{1}{2} v(\mathbf{x}) + L[v](\mathbf{x}) + \frac{i}{k} \frac{\partial}{\partial \mathbf{n}_{\mathbf{x}}} L[v](\mathbf{x}), \quad \mathbf{x} \in \Gamma \quad (2)$$

with

$$\begin{aligned} L[\phi](\mathbf{x}) &:= \int_{\Gamma} H(\mathbf{x}, \mathbf{y}) \phi(\mathbf{y}) d\mathbf{y}, \\ L[v](\mathbf{x}) &:= \int_{\Gamma} G(\mathbf{x}, \mathbf{y}) v(\mathbf{y}) d\mathbf{y}. \end{aligned} \quad (3)$$

$G(\mathbf{x}, \mathbf{y})$ and $H(\mathbf{x}, \mathbf{y})$ are the Green's function of the Helmholtz equation and its derivative with respect to the normal vector, respectively [3].

2.4.2. Stage 1: Building

To discretize Eq. (2), the surface Γ is represented by a mesh consisting of a set of triangular elements Γ_j . In our approach collocation with constant elements is used [16] and Eq. 2 is transformed into a system of equations:

$$\mathbf{A}\phi = \mathbf{b}, \quad (4)$$

where ϕ is the vector of unknown velocity potentials $\phi_i = \phi(\mathbf{x}_j)$ at the collocation nodes \mathbf{x}_j (in our case the midpoints of the elements). The system matrix \mathbf{A} is given by

$$\begin{aligned} A_{ij} = & -\frac{1}{2}\delta_{i,j} + \int_{\Gamma_j} H(\mathbf{x}_i, \mathbf{y}) d\mathbf{y} + \\ & + \frac{i}{k} \int_{\Gamma_j} \frac{\partial}{\partial \mathbf{n}_{\mathbf{x}}} H(\mathbf{x}_i, \mathbf{y}) d\mathbf{y} \end{aligned} \quad (5)$$

and

$$\begin{aligned} b_i = & \frac{i}{2k} v_i + \sum_j v_j \int_{\Gamma_j} G(\mathbf{x}_i, \mathbf{y}) d\mathbf{y} \\ & + \frac{i}{k} \sum_j v_j \int_{\Gamma_j} \frac{\partial}{\partial \mathbf{n}_{\mathbf{x}}} G(\mathbf{x}_i, \mathbf{y}) d\mathbf{y}. \end{aligned} \quad (6)$$

δ_{ij} denotes the kronecker delta symbol, v_i is the particle velocity at (constant) element Γ_i , and therefore also particle velocity at the collocation node. For most integrals standard Gauss-Legendre quadrature can be used, however when $i = j$ the integrands in Eq. (5) become (hyper)singular and special quadrature methods have to be applied. For a detailed description refer to [17, 4].

2.4.3. Stage 2: Solving

The integrals in Eqs. (5) and (6) are discretized and the solution of the linear system is calculated by using an iterative solver. The ML-FMM is used to speed up the matrix-vector multiplications needed for the iterative solver.

2.4.4. Stage 3: Post-Processing

Given the solution at the boundary ϕ , the sound pressure $p(\mathbf{x}) = i\rho\omega\phi(\mathbf{x})$ at any point \mathbf{x} in the exterior domain Ω_e can be determined using

$$\phi(\mathbf{x}) = \sum_j \int_{\Gamma_j} H(\mathbf{x}, \mathbf{y}) d\mathbf{y} \phi_j - \int_{\Gamma_j} G(\mathbf{x}, \mathbf{y}) d\mathbf{y} v_j. \quad (7)$$

Please note, that for the simulations the head is assumed to be sound hard, and the HRTFs are calculated by placing a sound source inside the ear using a velocity boundary condition (c.f. [3]), almost all v_j will be zero except for a few elements at the closure at the ear canal. HRTFs are evaluated for the same points as in 2.1 and for an ico-sphere with 2562 nodes.

2.5. Range Extrapolation

Although Eq. (7) can be represented by a simple matrix vector multiplication, this multiplication can become very time consuming. The FMM is not efficient if the evaluation nodes are far from the head (i.e. if the HRTFs for sources far from the head have to be calculated), thus it is proposed to calculate the HRTFs on a sphere around the head with small radius and extrapolate the results to larger spheres using spherical harmonics [11, 7].

For a fixed distance, the pressure $p(\mathbf{x})$ and thus the HRTFs (for arbitrary directions) can be seen as a function defined on the sphere. Under the assumption that inward traveling wave components can be neglected, the acoustic pressure at a point $\mathbf{x} = (r, \theta, \phi)$ on a sphere with radius r can be expanded using spherical harmonics

$$p(r, \theta, \phi, k) = \sum_{mn=0}^{\infty} a_{mn}(k, r) Y_n^m(\theta, \phi), \quad (8)$$

where

$$Y_n^m(\theta, \phi) = (-1)^m \sqrt{\frac{(2n+1)(n-|m|)!}{2\pi(n+|m|)!}} \cdot P_n^{|m|}(\cos(\theta)) e^{im\phi}. \quad (9)$$

$P_n^{|m|}$ is given by the associated Legendre polynomial and $a_{mn} = b_{nm}(k)h_n(kr)$ with $h_n(x)$ being the spherical Hankel function of order n and $\sum_{nm=0}^{\infty} = \sum_{n=0}^{\infty} \sum_{m=-n}^n$.

The unknown coefficients a_{nm} can either be determined using the orthogonality of the Y_n^m with respect to the inner product on the sphere or by using a least square solution of Eq. (8) at given sample points, where the first sum is truncated at a maximum order N :

$$\mathbf{a} = \mathbf{Y}^\dagger \mathbf{p}, \quad (10)$$

where \mathbf{a} is a vector containing the expansion coefficients $a_{mn}(k, r)$, \mathbf{p} is a vector containing the pressure at the sample points $x_i = (\theta_i, \phi_i)$ on the sphere and the entries of the matrix \mathbf{Y} are given by the spherical harmonics $Y_{mn}(\theta_i, \phi_i)$ evaluated at the sample points x_i . For our numerical experiments the maximum spherical harmonics order was set to $N = 42$ and the sphere was discretized by an ico-sphere with 2562 nodes.

Once the expansion coefficients $a_{mn}(k, r)$ are determined, the pressure at a sphere with radius R is given as

$$p(R, \theta_i, \phi_i, k) = \sum_{mn=0}^N \frac{a_{mn} h_n(kR)}{h_n(kr)} Y_{nm}(\theta_i, \phi_i) \quad (11)$$

2.6. Evaluation

Numerically simulated and acoustically measured HRTFs were compared and evaluated using two models. First, spectral features were analyzed with a model of sound-localization performance in sagittal-planes [18] implemented as *baumgartner2013* in the auditory modeling toolbox² (AMT). This model quantifies the predicted individual localization performance by the local polar RMS error (PE) in degrees and by the quadrant error rate (QE) in percent - common error measurements in localization experiments. Second, temporal features were analyzed by a direction-continuous model of the time-of-arrival [19] implemented as *ziegelwanger2014* in the AMT. The temporal features are quantified as a modeled head radius.

3. Results

Fig. 3 shows amplitude spectra in the median-plane (left panels) and energy time curves (ETCs) in the horizontal-plane (right panels) of listener NH5 for the reference mesh, a uniformly discretized mesh with 5 mm AEL and for a non-uniformly discretized mesh with edge lengths ranging from 1 to 15 mm (far-field, near-field and range-extrapolated). Note, that the uniformly and the non-uniformly discretized meshes had almost the same number of DOFs.

Tab. I shows the predicted individual localization performance in terms of PE and QE for the acoustic measurement, all numeric simulations and for non-individual HRTFs of the KEMAR mannequin³.

Tab. II shows the total computation time in hours (obtained on a single core of an Intel i7 CPU) and the relative computation time of the three stages in the BEM for simulating HRTFs of the reference mesh and a non-uniformly discretized mesh with edge lengths ranging from 1 to 15 mm in the far-field and the near-field.

4. Discussion

4.1. Uniform Meshes

For the uniform discretization conditions the AEL was systematically increased from 2 to 5 mm. The number of elements in the mesh decreased on average from 110 159 to 18 142 elements. The localization performance decreased with increasing AEL for all subjects. The modeled head-radius did not vary significantly over conditions. The localization performance for the 5 mm AEL mesh was in the range of

² Available from <http://amtoolbox.sourceforge.net> (date last viewed May 27, 2014)

³ Available from <http://interface.cipic.ucdavis.edu/sound/hrtf.html> (date last viewed May 27, 2014)

Table I. Mesh parameters (min. and max. edge length and number of elements in the meshes) and evaluation results from the localization model for the measured and simulated HRTFs and for non-individual KEMAR HRTFs.

Distance	Edge Length			DOF (mean)	NH5			NH130			NH131		
	Min. (mm)	Max. (mm)	Mean (mm)		PE (deg)	QE (%)	r (mm)	PE (deg)	QE (%)	r (mm)	PE (deg)	QE (%)	r (mm)
Individual Acoustic Measurement					29.4	13.1	92.1	25.9	10.8	92.2	31.3	13.6	94.7
Far-Field	1	2.5	1.7	110159	30.4	12.1	90.8	29.5	14.7	89.4	30.7	14.3	94.6
	2	2	2	106526	30.4	12.4	91.4	29.1	15.6	89.5	31	15	94.6
	3	3	3	53671	32.1	13.7	91.6	29.4	15.7	89.4	32.8	20	94.6
	4	4	4	28188	34.1	13.9	91.7	32.3	18.3	90.2	34.5	22.7	94.5
	5	5	5	18142	35.6	16.2	91.1	35.6	24.2	89.8	38.3	25.7	95.1
	2	4	3.7	26505	31.8	14.4	90.9	28.3	15.5	88.3	32.5	16.7	93.4
	2	8	4.3	18645	31.9	14.1	91.5	28.3	15.6	88.4	32.7	17.4	93.4
	2	15	4.7	13754	31.8	14	91.2	28.3	15.3	88.4	32.7	16.9	93.3
	1	15	3.9	13757	31.6	10.1	92.4	29.8	15.5	88.9	32.3	10.3	94.8
Near-Field	1	2.5	1.7	110159	31.8	9.9	93.7	30.1	13.5	91.2	31.3	10	96.6
	1	15	3.9	13757	31.6	10.1	92.8	29.8	15.5	91.5	32.3	10.3	96.3
Non-Individual KEMAR HRTFs					36.7	19.7	92.3	36.8	26.8	92.3	37.3	23.1	92.3

localization performance with non-individual HRTFs (KEMAR). Thus, a uniform discretization is not applicable to reduce the computational effort.

4.2. Non-Uniform Meshes

For the non-uniform discretization conditions the maximal edge length was varied from 2 to 15 mm. The number of elements in the mesh decreased from 110 159 to 13 754 elements. The localization performance was almost constant for all conditions and was in the same range as results for the uniform meshes with 2 and 3 mm AEL. The predicted sound localization performance, thus, seems to depend on the minimal edge length at the ipsi-lateral pinna and seems to be not affected by the increased edge length on the contra-lateral side. Based on these results, we also simulated HRTFs for a non-uniform discretization with edge lengths ranging from 1 to 15 mm. Here the localization performance was in the range of the reference mesh. However, the computation time was decreased from 83.9 to 14.7 hours (compare Tab. II).

4.3. Near-field Simulation

The simulation of HRTFs in the near-field of the head aims in reducing the numerical costs of the post-processing stage in the BEM. The effect for the simulation for the reference mesh is rather small. However, the near-field simulation reduces the computation time by almost 30 percent for the non-uniformly discretized meshes. The combination of non-uniform discretization and near-field simulation reduces the total computation time by a factor of 8. The localization performance for these HRTFs is in the range of the localization performance evaluated for the reference mesh (compare Tab. I).

Table II. Total simulation time and the relative computation time for the three stages in the simulation process in the far-field and the near-field and for uniformly and non-uniformly discretized meshes (averaged over listeners).

Computation-time	Far-field		Near-field	
	Unif.	N.-Unif.	Unif.	N.-Unif.
Building [%]	33.3	44.7	35.4	61.2
Solving [%]	56.5	17.9	60.2	24.5
Post-Proc. [%]	10.2	37.3	4.3	14.3
Total [h]	83.9	14.7	78.7	10.8

5. Conclusions

Our results show two advantages of a non-uniform discretization over the uniform discretization. First, the total computation time has been reduced by a factor of 6 without substantially affecting the modeled localization performance and broadband temporal features of HRTFs. Second, the near-field simulation further reduced the computation time by approximately 30% when applied on the non-uniformly discretized meshes (but not on the uniformly discretized meshes). The combination of a non-uniform mesh, the near-field approach, and four simultaneously processing CPUs allows to simulate HRTFs within a similar duration required for a corresponding acoustic measurement.

Acknowledgement

This study was supported by the Austrian Science Fund (FWF, projects P 24124-N13).

References

- [1] J. Blauert: Spatial hearing: the psychophysics of human sound localization. MIT Press, Cambridge, Mass, 1997.
- [2] P. Majdak, P. Balazs, B. Laback: Multiple exponential sweep method for fast measurement of head-

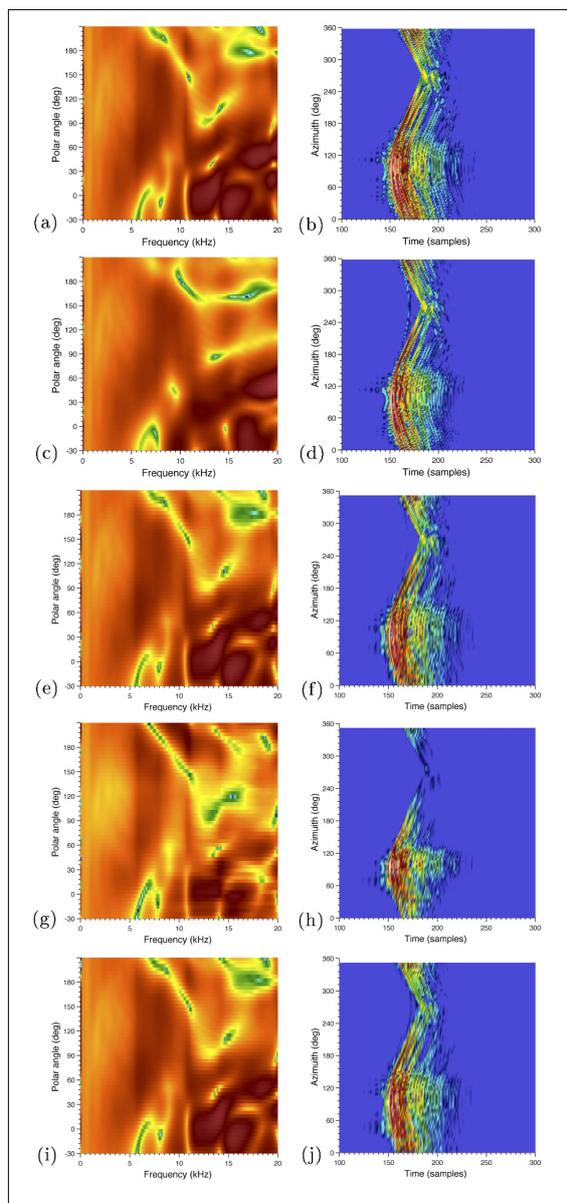


Figure 3. Median-plane amplitude spectra (a, c, e, g, i) and horizontal-plane energy time curves (b, d, f, h, j) for the reference mesh (a, b), uniformly discretized mesh with 15 mm AEL (c, d), non-uniformly discretized mesh with 1-15 mm AEL simulated in the far-field (e, f), near-field (g, h) and near-field with range-extrapolation (i, j).

- related transfer functions. *J. Audio Eng. Soc.* **55** (2007) 623–637.
- [3] W. Kreuzer, P. Majdak, Z. Chen: Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range. *J. Acoust. Soc. Am.* **126** (2009) 1280–1290.
- [4] Z.-S. Chen, H. Waubke, W. Kreuzer: A formulation of the fast multipole boundary element method fm/bem for acoustic radiation and scattering from

three-dimensional structures. *J. Comput. Acoust.* **16** (2008) 303–320.

- [5] S. Marburg: Six boundary elements per wavelength: Is that enough? *J. Comput. Acoust.* **10** (2002) 25–51.
- [6] H. Ziegelwanger, A. Reichinger, P. Majdak: Calculation of listener-specific head-related transfer functions: Effect of mesh quality. Proceedings of the 21st ICA, Montreal, Canada, 2013, ASA, 050017.
- [7] N. A. Gumerov, A. E. O'Donovan, R. Duraiswami, D. N. Zotkin: Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation. *J. Acoust. Soc. Am.* **127** (2010) 370–386.
- [8] J. T. Chen, K. H. Chen, C. T. Chen: Adaptive boundary element method of time-harmonic exterior acoustics in two dimensions. *Comput. Method Appl. M.* **191** (2002) 3331–3345.
- [9] H. Ziegelwanger, P. Majdak, W. Kreuzer: Ungleichförmige geometriediskretisierung für die numerische berechnung von außenohrübertragungsfunktionen. Fortschritte der Akustik - DAGA 2014, Oldenburg, DE, 2014.
- [10] H. Ziegelwanger, P. Majdak, W. Kreuzer: Non-uniform sampling of geometry for the numeric simulation of head-related transfer functions. Proceedings of the 21st International Congress of Sound and Vibration, Beijing, CN, 2014.
- [11] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, M. Noisternig: Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition. *Acta Acust. Acust.* **98** (2012) 72–82.
- [12] P. Majdak, M. J. Goupell, B. Laback: 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. *Atten. Percept. Psychophys.* **72** (2010) 454–69.
- [13] M. Botsch, L. Kobbelt: A remeshing approach to multiresolution modeling. Proceedings of the 2004 Eurographics/ACM SIGGRAPHs, New York, NY, USA, 2004, ACM, 185–192.
- [14] J. Möbius, L. Kobbelt: OpenFlipper: an open source geometry processing and rendering framework. Proceedings of the 7th International Conference on Curves and Surfaces, Berlin, Heidelberg, 2012, Springer-Verlag, 488–500.
- [15] A. J. Burton, G. F. Miller: The application of integral equation methods to the numerical solution of some exterior boundary-value problems. Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences **323** (Juni 1971) 201–210.
- [16] P. Hunter, A. Pullan: FEM/BEM notes. Tech. Rept. 2002.
- [17] M. Fischer, U. Gauger, L. Gaul: A multipole galerkin boundary element method for acoustics. *Engineering Analysis with Boundary Elements* **28** (2004) 155–162.
- [18] R. Baumgartner, P. Majdak, B. Laback: Assessment of sagittal-plane sound localization performance in spatial-audio applications. – In: *The Technology of Binaural Listening*. Springer, Berlin, Germany, 2013.
- [19] H. Ziegelwanger, P. Majdak: Modeling the direction-continuous time-of-arrival in head-related transfer functions. *J. Acoust. Soc. Am.* **135** (2014) 1278–1293.

Chapter 9

Concluding remarks

In this PhD project, requirements for the numerical calculation of listener-specific HRTFs were evaluated based on perception-based model predictions and sound-localization experiments. Two models were used, i.e., an existing sagittal plane sound-localization model and a TOA model. The development, implementation, and evaluation of the TOA model was part of the PhD project. Its evaluation showed that the TOA in HRTFs and, thus, the ITD can be described by a geometric spherical-head model. The main parameters to describe the TOA were the listener's head radius, the head position, and the position of the ear canals. By modeling the head position, the determination of a listener's head radius was robust against head positioning errors in the simulation or measurement of HRTFs.

The use of model predictions in advance of sound-localization experiments enabled the evaluation of different parameters effecting numerically calculated HRTFs and the evaluation of a large number of conditions. Following requirements for the numerical calculation of listener-specific HRTFs were identified:

1. Modeling the microphone by a single triangular element yielded a sound-localization performance similar to that which could be observed when subjects localize sound sources with acoustically measured HRTFs, as long as the position of the virtual microphone was located at the area of the blocked ear canal. Apart from the blocked ear canal, the modeled sound-localization performance degraded with the distance of the virtual microphone to the ear-canal. Increasing the area of the virtual microphone by averaging the sound pressure over numerous elements did not improve the modeled sound-localization performance.
2. When a uniform mesh was used for the numerical HRTF calculation, an average edge length of 2 mm was required to retrieve HRTFs which yield a sound-localization

performance similar to that can be observed when subjects localize sound sources with acoustically measured HRTFs.

3. A refinement of a mesh does not improve the results of the numerical HRTF calculation, if the resolution of the initial mesh does not meet the requirement of an AEL of 2 mm. The geometrical error is the dominant factor affecting the calculation results.
4. A uniform geometry discretization or a homogenous vertex distribution in the mesh was not mandatory for accurate results. Further, a non-uniform mesh generated by a-priori mesh grading reduced the computation time while preserving a numerical and perceptual error as low as when a uniform mesh with an AEL of 2 mm for the calculation was used.
5. HRTFs were efficiently calculated in the near-field. When a non-uniform mesh was used for the numerical calculation, the combination of an evaluation grid on the bounding sphere of the the head mesh and range extrapolation using spherical harmonics approximation further reduced the calculation time.

The numerical calculation of listener-specific HRTFs in this PhD project based on laser scans of heads as well as high-energy CT scans of pinna silicone imprints. The calculations were performed on high-end personal computers. Although the underlying geometry acquisition technique is not applicable in consumer products, this PhD thesis demonstrates that the numerical methods are capable of calculating listener-specific HRTFs that yield the same localization performance as observed when listeners localize sound sources with acoustically measured HRTFs. A-priori mesh grading and near-field calculation reduce the computational effort in the HRTF calculation process significantly and make the calculation applicable on smartphones for instance. For an end-user application on smartphones, however, an accurate and simple acquisition of a listener's pinnae is essential. Thus, the development and improvement of simple geometry acquisition techniques is necessary. Photogrammetric reconstruction shows a high potential for the acquisition of a listener's pinnae geometry (Reichinger *et al.*, 2013). While photogrammetric reconstruction was successfully applied to calculate low-frequency HRTFs of animals (Dellepiane *et al.*, 2008), further investigations and improvements are required to achieve an adequate accuracy when capturing the geometry of a human pinna, since pinna meshes acquired by photogrammetric reconstruction suffer from data gaps in the retrieved point clouds, e.g., in regions behind the helix or in the concha. Such data gaps might be reduced by the use of stereo image data of multiple cameras (Harder *et al.*, 2013, 2016) or by correcting point cloud data with a parametric or skeleton model (Amenta *et al.*, 1998)

of the human pinna. Also pinna feature extraction from images (Spagnol *et al.*, 2013, 2014; Liu and Zhong, 2015; Torres-Gallegos *et al.*, 2015) might be used to improve the final mesh accuracy, e.g., by extracting edges of the helix, the antihelix or the concha. Further reduction of the computational effort might be achieved by a frequency dependent a-priori mesh grading, where the number of elements in the mesh is optimized for the actual frequency in the calculation.

The software used in this PhD project was published open-source. The software package used for the numerical calculation of HRTFs was published as *Mesh2HRTF*¹. For compatibility with existing HRTF databases, *Mesh2HRTF* stores HRTF sets as SOFA files, i.e., the **S**patially **O**riented **F**ormat for **A**coustics² which is the official standard of the Audio Engineering Society for storing HRTFs (Majdak *et al.*, 2013). The TOA model was published as *ziegelwanger2014* in the Auditory Modeling Toolbox³ (AMT, Søndergaard and Majdak, 2013). In future work, *Mesh2HRTF* can be used to investigate the range dependence of HRTFs and in a more perceptual manner the corresponding distance perception and externalization. The TOA model can be extended to model the ILD in HRTFs (Spagnol and Avanzini, 2015). The release of *Mesh2HRTF* and *ziegelwanger2014* is intended to stimulate further research in the field of numerical HRTF calculation, the approximation and modeling of HRTFs, and binaural virtual acoustics in general.

¹Mesh2HRTF: available from <http://mesh2hrtf.sourceforge.net>

²SOFA: API available from <http://www.sofaconventions.org>

³AMT: available from <http://amttoolbox.sourceforge.net>

Bibliography

Algazi, V. R., Avendano, C., and Duda, R. O. (2001). “Elevation localization and head-related transfer function analysis at low frequencies”, *The Journal of the Acoustical Society of America* **109**, 1110–1122.

URL <http://dx.doi.org/10.1121/1.1349185>

Amenta, N., Bern, M., and Eppstein, D. (1998). “The Crust and the β -Skeleton: Combinatorial Curve Reconstruction”, *Graphical Models and Image Processing* **60**, 125–135.

URL <http://dx.doi.org/10.1006/gmip.1998.0465>

Baumgartner, R., Majdak, P., and Laback, B. (2013). “Assessment of Sagittal-Plane Sound Localization Performance in Spatial-Audio Applications”, in *The Technology of Binaural Listening*, 93–119 (Springer, Berlin, DE).

URL http://dx.doi.org/10.1007/978-3-642-37762-4_4

Baumgartner, R., Majdak, P., and Laback, B. (2014). “Modeling sound-source localization in sagittal planes for human listeners”, *The Journal of the Acoustical Society of America* **136**, 791–802.

URL <http://dx.doi.org/10.1121/1.4887447>

Blauert, J. (1997). *Spatial hearing: the psychophysics of human sound localization* (MIT Press, Cambridge, Mass).

Botsch, M. and Kobbelt, L. (2004). “A Remeshing Approach to Multiresolution Modeling”, in *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geom-*

etry Processing, SGP '04, 185–192 (ACM, New York, NY, USA).

URL <http://dx.doi.org/10.1145/1057432.1057457>

Brebbia, C. and Dominguez, J. (1977). “Boundary element methods for potential problems”, *Applied Mathematical Modelling* **1**, 372 – 378.

URL [http://dx.doi.org/10.1016/0307-904X\(77\)90046-4](http://dx.doi.org/10.1016/0307-904X(77)90046-4)

Bronkhorst, A. W. (1995). “Localization of real and virtual sound sources”, *The Journal of the Acoustical Society of America* **98**, 2542–2553.

URL <http://dx.doi.org/10.1121/1.413219>

Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., and Thompson, M. R. (2014). *Magnetic Resonance Imaging: Physical Principles and Sequence Design* (John Wiley & Sons Ltd).

URL <http://dx.doi.org/10.1002/9781118633953.ch1>

Carlile, S., Delaney, S., and Corderoy, A. (1999). “The localisation of spectrally restricted sounds by human listeners”, *Hearing Research* **128**, 175–189.

URL [http://dx.doi.org/10.1016/S0378-5955\(98\)00205-6](http://dx.doi.org/10.1016/S0378-5955(98)00205-6)

Cessenat, O. and Despres, B. (1998). “Application of an ultra weak variational formulation of elliptic pdes to the two-dimensional helmholtz problem”, *SIAM Journal on Numerical Analysis* **35**, 255–299.

URL <http://dx.doi.org/10.1137/S0036142995285873>

Chen, J. T., Chen, K. H., and Chen, C. T. (2002). “Adaptive boundary element method of time-harmonic exterior acoustics in two dimensions”, *Computer Methods in Applied Mechanics and Engineering* **191**, 3331–3345.

URL [http://dx.doi.org/10.1016/S0045-7825\(02\)00214-1](http://dx.doi.org/10.1016/S0045-7825(02)00214-1)

Chen, Z.-S., Waubke, H., and Kreuzer, W. (2008). “A formulation of the fast multipole boundary element method (FMBEM) for acoustic radiation and scattering from three-

dimensional structures”, *Journal of Computational Acoustics* **16**, 303–320.

URL <http://dx.doi.org/10.1142/S0218396X08003725>

Ciskowski, R. and Brebbia, C. (1991). *Boundary Element Methods in Acoustics* (Springer Netherlands).

Coleman, P. D. (1963). “An analysis of cues to auditory depth perception in free space”, *Psychological Bulletin* **60**, 302–315.

URL <http://dx.doi.org/10.1037/h0045716>

Dellepiane, M., Pietroni, N., Tsingos, T., Asselot, M., and Scopigno, R. (2008). “Reconstructing head models from photographs for individualized 3d-audio processing”, *Comput. Graph. Forum* 1719–1727.

URL <http://dx.doi.org/10.1111/j.1467-8659.2008.01316.x>

Dietrich, P., Masiero, B., and Vorländer, M. (2013). “On the optimization of the multiple exponential sweep method”, *J. Audio Eng. Soc* **61**, 113–124.

Goldstein, C. I. (1982). “The finite element method with non-uniform mesh sizes applied to the exterior Helmholtz problem”, *Numerische Mathematik* **38**, 61–82.

URL <http://dx.doi.org/10.1007/BF01395809>

Greengard, L. and Rokhlin, V. (1987). “A fast algorithm for particle simulations”, *Journal of Computational Physics* **73**, 325 – 348.

URL [http://dx.doi.org/10.1016/0021-9991\(87\)90140-9](http://dx.doi.org/10.1016/0021-9991(87)90140-9)

Gumerov, N. A., O’Donovan, A. E., Duraiswami, R., and Zotkin, D. N. (2010). “Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation”, *The Journal of the Acoustical Society of America* **127**, 370–386.

URL <http://dx.doi.org/10.1121/1.3257598>

Harder, S., Paulsen, R. R., Larsen, M., and Laugesen, S. (2013). “A three dimensional children head database for acoustical research and development”, Proceedings of Meetings on Acoustics **19**, 050013.

URL <http://dx.doi.org/10.1121/1.4800159>

Harder, S., Paulsen, R. R., Larsen, M., Laugesen, S., Mihocic, M., and Majdak, P. (2016). “A framework for geometry acquisition, 3-D printing, simulation, and measurement of head-related transfer functions with a focus on hearing-assistive devices”, Computer-Aided Design **75–76**, 39–46.

URL <http://dx.doi.org/10.1016/j.cad.2016.02.006>

Hartmann, W. M. and Wittenberg, A. (1996). “On the externalization of sound images”, The Journal of the Acoustical Society of America **99**, 3678–3688.

URL <http://dx.doi.org/10.1121/1.414965>

Heinrich, W., Beilenhoff, K., Mezzanotte, P., and Roselli, L. (1996). “Optimum mesh grading for finite-difference method”, IEEE Transactions on Microwave Theory and Techniques **44**, 1569–1574.

URL <http://dx.doi.org/10.1109/22.536606>

Heritage, G. L. and Large, A. R. G. (2009). *Laser Scanning for the Environmental Sciences* (Wiley-Blackwell).

URL <http://dx.doi.org/10.1002/9781444311952>

Herman, G. T. (2009). *Fundamentals of Computerized Tomography* (Springer-Verlag London).

URL <http://dx.doi.org/10.1007/978-1-84628-723-7>

Huttunen, T., Seppälä, E. T., Kirkeby, O., Kärkkäinen, A., and Kärkkäinen, L. (2007). “Simulation of the transfer function for a head-and-torso model over the entire audible frequency range”, Journal of Computational Acoustics **15**, 429–448.

URL <http://dx.doi.org/10.1142/S0218396X07003469>

Jin, C. T., Guillon, P., Epain, N., Zolfaghari, R., van Schaik, A., Tew, A. I., Hetherington, C., and Thorpe, J. (2014). “Creating the Sydney York Morphological and Acoustic Recordings of Ears Database”, *IEEE Transactions on Multimedia* **16**, 37–46.

URL <http://dx.doi.org/10.1109/TMM.2013.2282134>

Kahana, Y. and Nelson, P. A. (2006). “Numerical modelling of the spatial acoustic response of the human pinna”, *Journal of Sound and Vibration* **292**, 148–178.

URL <http://dx.doi.org/10.1016/j.jsv.2005.07.048>

Kahana, Y. and Nelson, P. A. (2007). “Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models”, *Journal of Sound and Vibration* **300**, 552–579.

URL <http://dx.doi.org/10.1016/j.jsv.2006.06.079>

Katz, B. F. G. (2001a). “Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation”, *The Journal of the Acoustical Society of America* **110**, 2440–2448.

URL <http://dx.doi.org/10.1121/1.1412440>

Katz, B. F. G. (2001b). “Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements”, *The Journal of the Acoustical Society of America* **110**, 2449–2455.

URL <http://dx.doi.org/10.1121/1.1412441>

Kreuzer, W., Majdak, P., and Chen, Z. (2009). “Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range”, *The Journal of the Acoustical Society of America* **126**, 1280–1290.

URL <http://dx.doi.org/10.1121/1.3177264>

Langendijk, E. H. A. and Bronkhorst, A. W. (2002). “Contribution of spectral cues to human sound localization”, *The Journal of the Acoustical Society of America* **112**,

1583–1596.

URL <http://dx.doi.org/10.1121/1.1501901>

Langer, U., Mantzaflaris, A., Moore, S. E., and Touloupoulos, I. (2015). “Mesh grading in isogeometric analysis”, *Computers & Mathematics with Applications* **70**, 1685 – 1700.

URL <http://dx.doi.org/10.1016/j.camwa.2015.03.011>

Liang, M. T., Chen, J. T., and Yang, S. S. (1999). “Error estimation for boundary element method”, *Engineering Analysis with Boundary Elements* **23**, 257–265.

URL [http://dx.doi.org/10.1016/S0955-7997\(98\)00086-1](http://dx.doi.org/10.1016/S0955-7997(98)00086-1)

Liu, X., Zhong, X. (2015). “Extracting Anthropometric Parameters from a Scanned 3D-Head-Model”, in *Proceedings of the 2nd International Conference on Information Science and Control Engineering (ICISCE)*, 225–228.

URL <http://dx.doi.org/10.1109/ICISCE.2015.57>

Lu-Xingchang and Liu-Xianlin (2006). *Innovations in 3D Geo Information Systems*, chapter Reconstruction of 3D Model Based on Laser Scanning, 317–332 (Springer, Berlin, Heidelberg).

URL http://dx.doi.org/10.1007/978-3-540-36998-1_25

Ma, F., Wu, J. H., Huang, M., Zhang, W., Hou, W., and Bai, C. (2015). “Finite element determination of the head-related transfer function”, *Journal of Mechanics in Medicine and Biology* **15**, 1550066 (20 pages).

URL <http://dx.doi.org/10.1142/S0219519415500669>

Macpherson, E. A. and Middlebrooks, J. C. (2002). “Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited”, *The Journal of the Acoustical Society of America* **111**, 2219–2236.

URL <http://dx.doi.org/10.1121/1.1471898>

Majdak, P., Balazs, P., and Laback, B. (2007). “Multiple Exponential Sweep Method

-
- for Fast Measurement of Head-Related Transfer Functions”, *Journal of the Audio Engineering Society* **55**, 623–637.
- Majdak, P., Baumgartner, R., and Laback, B. (2014). “Acoustic and non-acoustic factors in modeling listener-specific performance of sagittal-plane sound localization”, *Frontiers in Psychology* **5**, 319.
URL <http://dx.doi.org/10.3389/fpsyg.2014.00319>
- Majdak, P., Goupell, M. J., and Laback, B. (2010). “3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training”, *Attention, perception & psychophysics* **72**, 454–469.
URL <http://dx.doi.org/10.3758/APP.72.2.454>
- Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., and Noisternig, M. (2013). “Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions”, in *Proceedings of the 134th Convention of the Audio Engineering Society*, Convention Paper 8880.
- Marburg, S. (2002). “Six boundary elements per wavelength: Is that enough?”, *Journal of Computational Acoustics* **10**, 25–51.
URL <http://dx.doi.org/10.1142/S0218396X02001401>
- Mehrgardt, S. and Mellert, V. (1977). “Transformation characteristics of the external human ear”, *The Journal of the Acoustical Society of America* **61**, 1567–1576.
URL <http://dx.doi.org/10.1121/1.381470>
- Mershon, D. H. and King, L. E. (1975). “Intensity and reverberation as factors in the auditory perception of egocentric distance”, *Perception & Psychophysics* **18**, 409–415.
URL <http://dx.doi.org/10.3758/BF03204113>
- Middlebrooks, J. C. (1999a). “Individual differences in external-ear transfer functions reduced by scaling in frequency”, *The Journal of the Acoustical Society of America*

106, 1480–1492.

URL <http://dx.doi.org/10.1121/1.427176>

Middlebrooks, J. C. (**1999b**). “Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency”, *The Journal of the Acoustical Society of America* **106**, 1493–1510.

URL <http://dx.doi.org/10.1121/1.427147>

Middlebrooks, J. C. and Green, D. M. (**1990**). “Directional dependence of interaural envelope delays”, *The Journal of the Acoustical Society of America* **87**, 2149–2162.

URL <http://dx.doi.org/10.1121/1.399183>

Møller, H. (**1992**). “Fundamentals of binaural technology”, *Applied Acoustics* **36**, 171–218.

URL [http://dx.doi.org/10.1016/0003-682X\(92\)90046-U](http://dx.doi.org/10.1016/0003-682X(92)90046-U)

Møller, H., Sørensen, M. F., Hammershøi, D., and Jensen, C. B. (**1995**). “Head-Related Transfer Functions of Human Subjects”, *Journal of the Audio Engineering Society* **43**, 300–321.

Pears, N., Liu, Y., and Bunting, P. (**2012**). *3D Imaging, Analysis and Applications* (Springer-Verlag London).

URL <http://dx.doi.org/10.1007/978-1-4471-4063-4>

Pollow, M., Nguyen, K.-V., Warusfel, O., Carpentier, T., Müller-Trapet, M., Vorländer, M., and Noisternig, M. (**2012**). “Calculation of Head-Related Transfer Functions for Arbitrary Field Points Using Spherical Harmonics Decomposition”, *Acta Acustica united with Acustica* **98**, 72–82.

URL <http://dx.doi.org/10.3813/AAA.918493>

Rébillat, M., Benichoux, V., Otani, M., Keriven, R., and Brette, R. (**2014**). “Estimation of the low-frequency components of the head-related transfer functions of animals from

photographs”, *The Journal of the Acoustical Society of America* **135**, 2534–2544.

URL <http://dx.doi.org/10.1121/1.4869087>

Reichinger, A., Majdak, P., R., S., and Maierhofer, S. (2013). “Evaluation of Methods for Optical 3-D Scanning of Human Pinnae”, in *Proceedings of the International Conference on 3D Vision - 3DV 2013*, 390–397.

URL <http://dx.doi.org/10.1109/3DV.2013.58>

Rui, Y., Yu, G., Xie, B., and Liu, Y. (2013). “Calculation of Individualized Near-Field Head-Related Transfer Function Database Using Boundary Element Method”, in *Proceedings of the 134th Convention of the Audio Engineering Society* (Rome, IT).

Søndergaard, P. and Majdak, P. (2013). “The Auditory Modeling Toolbox”, in *The Technology of Binaural Listening*, 33–56 (Springer, Berlin, DE).

URL http://dx.doi.org/10.1007/978-3-642-37762-4_2

Spagnol, S., Geronazzo, M., and Avanzini, F. (2013). “On the Relation Between Pinna Reflection Patterns and Head-Related Transfer Function Features”, *IEEE Transactions on Audio, Speech, and Language Processing* **21**, 508–519.

URL <http://dx.doi.org/10.1109/TASL.2012.2227730>

Spagnol, S., Geronazzo, M., Rocchesso, D., and Avanzini, F. (2014). “Synthetic individual binaural audio delivery by pinna image processing”, *International Journal of Pervasive Computing and Communications* **10**, 239–254.

URL <http://dx.doi.org/10.1108/IJPCC-06-2014-0035>

Spagnol, S., and Avanzini, F. (2015). “Anthropometric tuning of a spherical head model for binaural virtual acoustics based on interaural level differences”, in *Proceedings of the International Conference on Auditory Display - ICAD 2015*, 204–209.

Takemoto, H., Mokhtari, P., Kato, H., Nishimura, R., and Iida, K. (2012). “Mechanism for generating peaks and notches of head-related transfer functions in the median plane”,

The Journal of the Acoustical Society of America **132**, 3832–3841.

URL <http://dx.doi.org/10.1121/1.4765083>

Torres-Gallegos, E. A., Orduña-Bustamante, F., and Arámbula-Cosío, F. (2015). “Personalization of head-related transfer functions (hrtf) based on automatic photanthropometry and inference from a database”, Applied Acoustics **97**, 84 – 95.

URL <http://dx.doi.org/10.1016/j.apacoust.2015.04.009>

Walsh, T. and Demkowicz, L. (2003). “hp Boundary element modeling of the external human auditory system—goal-oriented adaptivity with multiple load vectors”, Computer Methods in Applied Mechanics and Engineering **192**, 125–146.

URL [http://dx.doi.org/10.1016/S0045-7825\(02\)00536-4](http://dx.doi.org/10.1016/S0045-7825(02)00536-4)

Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (1993). “Localization using nonindividualized head-related transfer functions”, The Journal of the Acoustical Society of America **94**, 111–123.

URL <http://dx.doi.org/10.1121/1.407089>

Wightman, F. L. and Kistler, D. J. (1989a). “Headphone simulation of free-field listening. I: Stimulus synthesis”, The Journal of the Acoustical Society of America **85**, 858–867.

URL <http://dx.doi.org/10.1121/1.397557>

Wightman, F. L. and Kistler, D. J. (1989b). “Headphone simulation of free-field listening. ii: Psychophysical validation”, The Journal of the Acoustical Society of America **85**, 868–878.

URL <http://dx.doi.org/10.1121/1.397558>

Xie, B. (2013). *Head-related transfer function and virtual auditory display*, 2 edition (J. Ross. Pub. Inc.).

Yee, K. (1966). “Numerical solution of initial boundary value problems involving maxwell’s equations in isotropic media”, IEEE Transactions on Antennas and Prop-

agation **14**, 302–307.

URL <http://dx.doi.org/10.1109/TAP.1966.1138693>

Zienkiewicz, O., Taylor, R., and Zhu, J., eds. (**2013**). *The Finite Element Method: Its Basis and Fundamentals*, seventh edition (Butterworth-Heinemann, Oxford).

URL <http://dx.doi.org/10.1016/B978-1-85617-633-0.00019-8>

**Numerical Calculation of Individual
Head-Related Transfer Functions of Human Listeners**

Harald Ziegelwanger

Doctoral School for Scholarly Doctoral Studies
University of Music and Performing Arts Graz